April 2014

Development and Validation of Accelerometer-Based Activity Classification Algorithms for Older Adults: A Machine Learning Approach

Jeffer Eidi Sasaki
University of Massachusetts - Amherst

DEVELOPMENT AND VALIDATION OF ACCELEROMETER-BASED ACTIVITY CLASSIFICATION ALGORITHMS FOR OLDER ADULTS: A MACHINE LEARNING APPROACH

A Dissertation Presented

by

JEFFER EIDI SASAKI

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2014

Department of Kinesiology

DEVELOPMENT AND VALIDATION OF ACCELEROMETER-BASED ACTIVITY CLASSIFICATION ALGORITHMS FOR OLDER ADULTS: A MACHINE LEARNING APPROACH

A Dissertation Presented

by

JEFFER EIDI SASAKI

Approved as to style and content by:

Patty S. Freedson, Chair

Jane Kent-Braun, Member

Richard Van Emmerik, Member

John Staudenmayer, External Member

Patty S. Freedson, Department Chair
Department of Kinesiology

ACKNOWLEDGMENTS

First and foremost, I would like to thank my adviser Dr. Patty Freedson for giving me the opportunity to be part of her lab. I still remember when I showed up on her door about 6 years ago and she brought me in to the lab and said: "*You can stay here as long as you want*". At that time I stayed for only 40 days, but it has now been 5 years since I officially became her student. I have a lot of stories from this period but two of them will always remind me of her. The first is, whenever she introduces me to someone she still asks the famous question: "Guess where Jeffer is from?". That has proven to be a legit question as most people answer China, Japan, Korean, but never Brazil. I get a kick out of this question every time. The second is she can now (once in a while) pronounce my last name correctly. However, I still prefer when she says "Jeffer Saski". Patty, during these 5 years, you have been a wonderful mentor and I have no words to express my gratitude for all you have done for me and every member of the lab. You taught me a lot and made sure I got the best learning experience. Thank you so much!

I also want to thank my committee for their helpful input and feedback. I want to start by thanking Jane. This project would certainly have been much harder without your valuable suggestions. During my proposal you said (more or less with these words) "Boy, with all that you are proposing you are going to stay here for your entire life!". I now appreciate your suggestion of cutting out part 3 of my dissertation project. I also appreciate your advice for going to the senior centers and the inclusion of the 400m-walk to my study protocol. As the readers will find in this dissertation, this recommendation was very important for demonstrating some applications of our proposed methods. I must also thank Dr. Erin Snook, who unfortunately is no longer a professor in our department but had contributed to this project. All of your comments and suggestions made this document better. I wish all the best to you in whatever path you decide to take in life. Dr. Van Emmerik, thanks for replacing Erin on my committee. I know it was a last minute request but I really appreciate having you on my committee. Perhaps the best class I've taken during my PhD was your KIN 797R "Control and Coordination of Human Movement Systems". At first I had no idea what everyone was talking about in class. However, by the end of the semester I was really enthusiastic by how much I learned. Thank you very much!

I want to specially thank Dr. Staudenmayer. John, without your help during the last two semesters I have no idea how I would have made it. I had too much data and no knowledge on programming in R. This pretty much meant I was not in the most favorable position! But for one year you came to the lab and sat with Amanda and I to teach something that was Greek to us. I have no words to say how much I appreciated it. I have learned a lot and your patience made all the difference in my learning experience. I will never be able to pay back.

For anyone who is or was a PhD student, you probably know that your labmates become your family. You see them more often than your own family. During the past five years, I was blessed to have so many awesome labmates. I want to thank Cheryl for all the good conversations and advice. You gave me a lot of encouragement when I first got here and that meant a lot to me. Sarah and Kate, thanks for helping me decide on a project and for helping me along the way. I always enjoyed hanging out with you and sharing good laughs. I miss it a lot. Sarah, thanks for making Ben. He is the most adorable child in the world. Thank God I gave you that 'Fertility Dolphin'. Kate, thanks for inviting Dinesh and I to the Cape. That was a really nice trip. I still can’t believe the size of fish we caught. You and your family were just wonderful to us. I will never forget it. I must also thank my friend Dinesh. Thanks for all your help with writing and for sharing your knowledge on activity monitors. I miss the short breaks we used to take for a nice ‘espresso’. I’m glad you became a coffee snobby! If I ever open a coffee shop in Brazil, I will make sure I have an espresso named after you ‘Dinesh espresso’! Mari, thanks for helping me with several aspects of this project. You were amazing. I really appreciated all your effort in recruiting participants for my study. Anna, thanks for the good energy you brought into the lab while you were our lab manager. I wish you all the best in Norway, Sweden or whichever Nordic country you’re at. I must also thank Cori for organizing some good sport events (e.g. soccer) that helped to keep me sane during data analysis. The roaches didn’t win the intramurals but we had too much fun!

Amanda, during these past 4 years you helped me so much. I have no idea how I would have made it through this program without your help. You were always the first to read my assignments and papers. You were always the first to offer me help and the person I always asked help to. I will miss working and hanging out with you. Thanks for sitting with me for one year to analyze my data. This dissertation was only possible because of your help. Many times when I was frustrated or tired you were there to help me out. I’m really thankful for having you as my friend. I’m pretty sure the future reserves great success for you. “Don’t give up on R codes”!

Albert, thanks for spending so much time reading and editing my final document. You have been looking at this document for 4 days. “Sensei, you need to rest up!” I certainly appreciate it. If you read this, remember… “Do the kick, do the kick!” I’m glad

you became part of our lab. You brought good energy and motivation! You are certainly the right person to work with Amanda. I’m pretty sure you guys will be a rock star team.

I want to thank my parents for supporting me in everything I do. I’m very lucky to have such good parents. Your love and dedication give me reason to keep moving forward. Thanks for everything!

Last but not least, I want to thank Camila for her patience. Five years ago you said I should pursuit my PhD. During this time, the distance was hard but you were always there to support me and cheer me up. Thank you!

NOTE TO READERS

During the course of this dissertation project, the dissertation committee recommended modifications to the original proposed study. The modifications that are relevant for the readers are explained below.

1) Modification to Hypothesis 2 from Study 1

Hypothesis 2 from Study 1 was modified to be more appropriate for the study. The modification was as follows:

<u>Original dissertation proposal</u>

Hypothesis 2: The activity classification algorithms will predict physical activity intensity (METs) more accurately than current linear regression models.

<u>Current document</u>

Hypothesis 2: The machine learning models will predict activity intensity with similar accuracy observed in previous studies using machine learning models in younger adults (RMSE: 0.43 – 1.22 MET)

<u>Justification</u>

Previous studies have shown that machine learning algorithms are superior to linear regression models in predicting activity intensity. Thus, it was deemed more appropriate to use results from these studies as reference values for our algorithms rather than comparing them to linear regression models that are less sophisticated.

2) Modification to Hypothesis 2 from Study 2

Hypothesis 2 was modified in Study 2 to be consistent with the main objective of the study (prediction of activity type in free-living older adults). The modification was as follows:

Original dissertation proposal

Hypothesis 2: The activity classification algorithms will assess time spent by free-living older adults in different activity intensity levels more accurately than accelerometer cut-point methods.

Current document

Hypothesis 2: Algorithms developed with free-living accelerometer data will classify activity type in free-living older adults more accurately than lab-based algorithms developed in Study 1

Justification

The dissertation committee suggested that the focus of the study should be to exclusively examine classification of activity type from accelerometer data in free-living older adults. The new hypothesis was based on previous studies reporting that machine learning algorithms trained on laboratory data are less accurate in classifying activity type in free-living conditions than algorithms trained on free-living accelerometer data.

3) Modifications to Methods Section

The methods described in this document have been revised and are not the same methods as described in the dissertation proposal. Modifications were completed according to suggestions and comments from committee members. The major modification was the exclusion of a third study, that proposed to examine the association of a 7-day activity monitoring period with physical function scores obtained from the 400 m walk and SF-36 questionnaire. To examine this association, we proposed to apply one of the machine learning algorithms (developed in this dissertation research) to assess habitual physical activity level of the participants. This study was excluded from the project because all members from the dissertation committee (Patty Freedson, Jane Kent-Braun, John Staudenmayer, and Erin Snook - former committee member who was later replaced by Richard Van Emmerik) agreed that the timeline for the project was not feasible.

4) Instruments and Measures described in the methods section (Chapter III) but not included in the dissertation studies (Chapter IV and V)

1) The Physical Activity Scale for the Elderly (PASE) and the SF-36 questionnaire were answered by the participants but were not used for the dissertation studies. Data from these instruments may be used in future studies.

2) As the focus of Study 2 became the classification of activity type, we did not use any measures of activity intensity that were obtained with DO and heart rate monitoring. The procedures for coding activity intensity using DO and for monitoring heart rate were described in the methods section (Chapter III) but were not included in the Study 2.

ABSTRACT

DEVELOPMENT AND VALIDATION OF ACCELEROMETER-BASED ACTIVITY CLASSIFICATION ALGORITHMS FOR OLDER ADULTS: A MACHINE LEARNING APPROACH

FEBRUARY 2014

JEFFER EIDI SASAKI, B.S., FEDERAL UNIVERSITY OF PARANA, BRAZIL

M.S., FEDERAL UNIVERSITY OF PARANA, BRAZIL

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Patty Freedson

Machine learning algorithms to classify activity type from wearable accelerometers are important to improve our understanding of the relationship between physical activity (PA) and risk for physical disability in older adults. Therefore, the main objective of this dissertation was to develop and evaluate machine learning algorithms to predict activity type and intensity in older adults from a commercially available accelerometer (ActiGraph GT3X+).

In Study 1, we developed machine learning algorithms to classify activity type and intensity from raw accelerometer data in older adults. Thirty-five older adults performed an activity routine comprised of different activities (5 min/activity) while wearing three ActiGraph GT3X+ activity monitors (dominant hip, wrist, and ankle) and a portable metabolic system. Accelerometer and steady-state metabolic data were used to develop artificial neural network, random forest, and support vector machine algorithms (ANN_{Lab}, RF_{Lab}, and SVM_{Lab}) to predict activity type and intensity in older adults using 20 s classification intervals. Classification accuracy of the models in detecting five activity categories ranged from 87% (ANN_{Lab} hip, RF_{Lab} hip, and SVM_{Lab} hip) to 96%

(SVM_{Lab} wrist). The biases and root mean squared errors (RMSE) for predicted METs ranged from -0.01 MET (RMSE: 0.54 MET) for the RF_{Lab} wrist algorithm to 0.02 MET (RMSE: 0.67 MET) for the ANN_{Lab} hip algorithm.

Study 2 evaluated the performance of the RF_{Lab} and SVM_{Lab} algorithms for predicting activity type in free-living conditions. Fifteen participants from Study 1 were observed for 2-3 h in their free-living environment while wearing three ActiGraph GT3X+ activity monitors (dominant hip, wrist, and ankle). The RF_{Lab} and $SVM_{Lab\ -}$ algorithms were applied to hip, wrist, and ankle accelerometer data to classify five activity categories. Direct observation of activity type and duration served as criterion measures to evaluate percent correct classification rates of the algorithms. Correct classification rates ranged from 49% (SVM_{Lab} hip, SVM_{Lab} wrist, and RF_{Lab} wrist) to 55% (SVM_{Lab} ankle). New RF and SVM algorithms were developed using free-living accelerometer data (RF_{FL} and SVM_{FL}) and different classification intervals were also applied. Correct classification of activity types for the RF_{FL} and SVM_{FL} ranged from 53% (SVM_{FL} wrist, 5 s classification intervals) to 71% (SVM_{FL} ankle, 30 s classification intervals). Overall correct classification rates of up to 76% (RF_{FL} hip and RF_{FL} ankle, 30 s classification intervals) were achieved when classifying only three activity categories.

Our machine learning algorithms accurately predict activity type from accelerometer data in older adults under ‘laboratory conditions’ but not in free-living conditions. We were able to improve free-living classification accuracy using algorithms developed under free-living conditions. Further refinement of the algorithms is required for achieving sufficient accuracy in classifying activity type in free-living older adults.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF TERMS AND ABBREVIATIONS

ANN – Artificial neural network, machine learning tool

Bias - Measurement bias is the average difference between predicted and criterion measures

CFS-DO – Continuous focal sampling direct observation

DO – Direct Observation, criterion measure for activity type

EE – Energy expenditure

MET – Metabolic equivalent defined as an oxygen consumption of 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$. This value approximates resting oxygen consumption (1)

Mobility disability – Inability to complete a 400-meter walk at a usual pace within 15 minutes (2)

MVPA – Moderate-to-vigorous physical activity, typically defined as an absolute activity intensity raging from 3-6 METs

Functional limitations – Restrictions in performing fundamental physical and metal actions used in daily life (3)

PA – Physical activity

PAEE – Physical activity energy expenditure

Disability – Functional limitation expressed in a social context (e.g., inability to shop independently) (4).

Physical Frailty – Losses of physiologic reserve that leads to increased risk of disability (5).

Physical Impairment – Dysfunction at the level of tissues, organs, and body systems (6).

RF – Random forest, machine learning tool

RMR – Resting metabolic rate

RPE – Rate of perceived exertion

SB – Sedentary behavior

SVM – Support vector machine, machine learning tool

CHAPTER I

INTRODUCTION

Statement of the Problem

Individuals aged 65 and over will represent approximately 20% of the United States population by 2030 (7). The gradual aging of the US population has led the research community to become increasingly concerned about how to maintain quality of life in older adults (8,9). Promotion of active lifestyles for the older is an effective strategy to maintain physical function and independence in older adults (10,11). To obtain a more comprehensive understanding of the relationship between physical activity, function and independence, it is essential to develop PA assessment methods that are specific to older adults as activities differ considerably across different age groups (12,13). It is also critical to consider and understand measurement accuracy when assessing PA in different age groups (14). Studies in older adults have failed to consider this by relying on the use of questionnaires to assess PA levels (11,15). These instruments are particularly problematic in accurately assessing physical activity in older adults. Questionnaires usually fail to produce accurate PA measures because they are highly dependent on cognitive function and individual perceptions (16). Older adults are more likely to suffer from cognitive impairment, a condition that has a direct negative impact on their ability to recall past information (17). In addition, inaccuracy is higher in this age group as they spend substantial time in low-intensity PA (e.g. household activities), a category that is difficult to recall (16,18).

Accelerometer-based PA monitors are a feasible option to overcome these limitations by allowing objective assessment of free-living PA (19). To date, several studies have used these monitors to assess PA in older adults (19–23). However, these studies only used cut-point methods derived from linear regression models to assess time spent in different intensity categories. An important limitation of these methods is their inaccuracy in assessing intensity of activities that produce acceleration signals non-linearly related to energy expenditure, for example, activities of daily living (ADLs) (24). This limitation may lead to considerable misclassification of activity level in those older adults who spend a greater portion of their time in ADLs (12).

It may valuable to use accelerometer-based PA monitors to assess activity type rather than just activity intensity in older adults (25–27). Assessment of energy expenditure (EE) and activity intensity has prevailed because of the vast literature examining the relationship between energy expenditure and health outcomes (28–31). However, new technologies (e.g., accelerometers, multi-sensor activity monitors) and analytic tools (e.g., computational methods, artificial intelligence) can now be applied to obtain different PA metrics, such as activity type (32). Studies in older adults have shown that specific types of structured activities or exercise programs lead to specific physiologic and performance adaptations (33–35). While these studies provide evidence of the influence of activity type on physical function, they only account for a very small period of an individual's day (e.g. 45-minutes, 1-hour). Activity monitors can be used to obtain information about activity performed outside of a formal exercise or PA program. They may be the key to better understanding the impact of free-living activity behavior on preservation of physical function in older adults.

Identifying activity type will also help to improve assessment of sedentary behavior (SB), a factor of fundamental importance in an increasingly inactive society

(36). Objective assessment of SB with hip-worn monitors has also been conducted using cut-points (37–39). This approach typically results in misclassification of ADLs and standing as SB (39). To examine the association between SB and physical function in older adults, researchers need accurate methods to quantify SB. Machine learning techniques, such as artificial neural networks (ANN), support vector machines (SVM), and decision tree classifiers, maybe a solution to better quantify SB from acceleration signals (32).

Recent studies have used machine learning methods to develop accelerometer-based activity classification algorithms and were successful in identifying different activity types (25,26,40). More specifically, these studies accurately identified sedentary, household, locomotion and sport activities. This is possible because different activities produce different signals containing patterns that can be identified even if activities are somewhat similar and performed by different individuals (26,32,41,42). These signal patterns can also be identified by machine learning techniques in order to improve PAEE prediction. For example, Rothney et al. (43) and Staudenmayer et al. (26) demonstrated that artificial neural networks (ANNs) predict PAEE more accurately than linear prediction models. In the study by Staudenmayer et al. (26), root mean square error (RMSE) was computed for different MET prediction methods. The lowest RMSEs for MET prediction (0.43 - 1.22 MET) were achieved by the ANN method in comparison to the RMSE values resulting from three linear regression models (0.73 – 2.09 METs).

Despite the promising results obtained in younger adults, classification algorithms developed for older adults have mostly focused on assessing different postures and detecting falls (44,45). Currently, the absence of a method that can identify diverse activities in older adults highlights the need to develop accelerometer-based activity classification algorithms for this age group. It is essential to identify types of activities (e.g. sedentary, locomotion, household, etc) commonly performed in naturalistic settings to better understand individuals' physical function within their environment. Currently, physical function is assessed using self-report instruments and/or physical performance tests (46–51). These methods are important in determining risk for physical disability; however, an accurate, objective, free-living PA and SB assessment method may permit early detection of older adults at risk for physical disability. Therefore, in this study, we propose to develop and test the validity of accelerometer-based activity classification algorithms for older adults in laboratory and free-living conditions.

Aims and Significance

Study 1: Development and Validation of Accelerometer-based Activity Classification Algorithms for Older Adults

The cut-point method has been the method of choice to translate accelerometer output into PA and SB metrics in older adults (19–23,37). However, the cut-point method is restricted to measuring time spent in SB and different PA intensity categories (24,52,53). There are potential applications of using accelerometer data to assess activity type in older adults. Machine learning techniques are alternative analytic tools that can be used for this purpose. These techniques have shown promising results in younger adults, with studies reporting accuracy rates greater than 80% for activity type recognition (25–27), including ambulatory, lifestyle, and sedentary activities. Currently, the potential for assessing activity type from accelerometer data has improved since the current technology of commercially available activity monitors allows for collection of raw acceleration signals (g) at high sampling rates (100+ Hz). Machine learning techniques are ideal for processing high sampling rate data as they can identify complex patterns contained within accelerometer data from different activities, and, also, because they improve performance when more data examples are provided.

Developing machine learning algorithms to classify activity type in older adults is important as these algorithms may allow for objective assessment of PA characteristics of interest, such as locomotion time and speed. Identifying these gait variables from free-living accelerometer data may help to identify those more likely to experience reductions in quality of life. Studies have shown that locomotion time and speed are related to survival time and risk for becoming frail later in life (54,55). Previous studies have predicted walking speed from accelerometers, but most of the methods were restricted to clinical settings, younger individuals, and prototype monitors (56–58). It is important to develop prediction models for processing data from commercially available activity monitors in older adults. This is key to increase activity monitoring in older adults given that large-scale studies rely on the use of commercially available activity monitors. In addition, examining optimal location for monitor placement is critical, as it may play a role on participants' compliance during studies monitoring PA for prolonged durations. Studies in younger adults have reported successful results classifying activity type from wrist- and ankle- worn activity monitors, suggesting that these body locations are alternatives for monitor placement (41,59).

Another reason for using machine learning techniques is to improve estimations of energy expenditure. Staudenmayer et al. (26) reported that an ANN produced bias[1] as

---

[1] Measurement bias is the average difference between predicted and actual values

low as 0.1 MET when predicting EE of different activities. In comparison to linear regression methods, the ANN produced the lowest bias for prediction of METs in their study. Considering the advantages of using machine learning techniques, the aim of this study is to develop algorithms for classifying activity type, intensity and locomotion speed from raw acceleration signals in older adults. Participants will visit the Physical Activity and Health Laboratory and performed different activities while wearing accelerometer-based PA monitors and a portable indirect calorimetry system. Machine learning techniques will be applied to raw acceleration data to develop algorithms to classify activity type and intensity, and estimate locomotion speed.

Hypotheses

H1.1: The machine learning models will accurately predict activity type in older adults (≥ 80% accuracy)

H1.2: The machine learning models will predict activity intensity with similar accuracy observed in previous studies using machine learning models in younger adults (bias < ±0.1 MET)

Exploratory analyses

1) Studies in younger adults suggest that placement of activity monitor on the wrist or ankle produce accurate or acceptable recognition for activity type (41,59). There is limited evidence as to what monitor placement produces the best recognition rate for activity type in older adults. Monitor placement is important for increasing compliance and also for reducing burden on the participant. Given this importance, we determined optimal location for activity monitor placement based on the results from this study. It

could not be hypothesized which placement would be superior for prediction of activity type and intensity since no similar studies have been conducted in older adults.

2) To date, few studies have estimated locomotion speed from acceleration signals. The few studies that have done this were conducted in younger individuals or using accelerometers designed for clinical settings (56–58). Objective assessment of locomotion speed in older adults is of major importance, as studies have reported associations of locomotion speed with survival time, risks for disability, and risks for becoming frail (54,60,61). Thus, we developed machine learning algorithms to predict locomotion speed from raw acceleration signals in older adults from this study. This was treated as an exploratory aim since no measurement error values could be found in the literature to serve as reference values for the present investigation. We could not hypothesize the degree of accuracy and measurement error to be expected from the models in predicting locomotion speed.

Study 2: Validation of the Accelerometer-Based Activity Classification Algorithms in Free-living Older Adults

Validity of prediction models in laboratory settings does not ensure validity in free-living conditions. Measurement accuracy may decrease substantially without the control of a laboratory setting (62–64). Thus, it is imperative to determine the accuracy of activity classification algorithms in free-living conditions before they are implemented in studies quantifying PA level and SB of older adults.

Accelerometer cut-point methods have been validated in free-living conditions using portable indirect calorimetry or doubly labeled water (53,65–68). These criterion measures cannot be used to validate activity type classification algorithms in free-living conditions because they do not provide information on activity type. Algorithms that estimate activity type have been tested in free-living conditions with user-annotated data, which rely on the user's ability in recording and coding activities they perform (25,63,64). Studies using this approach have indicated substantial reduction in activity type recognition when classification algorithms developed in laboratory were used in free-living settings. For example, Gyllensten and Bonomi (64) reported reductions in recognition rate of approximately 16 to 20%. Similarly, Ermes et al. (63) observed that algorithms trained with laboratory data were up to 17% less accurate in identifying activity type in free-living conditions compared to algorithms trained with both laboratory and free-living accelerometer data. Thus, significant reduction in recognition rate of activity type is expected when algorithms developed in laboratory are applied to free-living conditions.

While user-annotated data have allowed for testing activity type classification algorithms in free-living conditions, it is also possible to obtain criterion data for activity type using direct observation (DO). Direct observation requires systematic training of observers for recording and coding activities in a consistent way. Very few studies have used DO to test the validity of activity type classification algorithms in free-living conditions. In one of these studies, Foerster et al. (62) observed that correct recognition rate of an activity type classification algorithm was 33.3% lower in free-living conditions compared to laboratory conditions. It is likely that trained observers are more meticulous in coding activity than users. As a result, activity recognition rates obtained with DO are likely more accurate and reliable than those obtained with user-annotation and this should be considered in studies validating activity classification algorithms in free-living conditions. Therefore, the primary aim of this study is to evaluate the field validity of the activity classification algorithms (developed in Study 1) in assessing activity type in free-living older adults using DO as a criterion measure. Trained observers will follow participants in a normal situation and record the activities performed during a 2-3 h time block. These data will be used to assess the accuracy of the activity classification algorithms from Study 1 in classifying activity type in free-living older adults. In addition, activity data obtained with DO will be used to improve classification of both free-living PA and SB from acceleration signals. A study by Ermes et al. (63) demonstrated that activity classification algorithms trained with both free-living and laboratory accelerometer data demonstrated an improvement of approximately 12% in identifying free-living activity type compared to algorithms trained using only laboratory data. A secondary aim of this study, which is an exploratory aim, is to demonstrate how activity classification algorithms may be applied to evaluate associations of a specific PA characteristic (i.e., locomotion speed) with a physical function score (i.e., 400m walk). We will use algorithms developed in Study 1 to predict locomotion speed in free-living conditions. Correlation analysis will be applied to examine the association of these variables with speed in the 400 m walk.

Hypotheses

H2.1: The machine learning algorithms developed in Study 1 will classify activity type from accelerometer data in free-living older adults with similar accuracy as previous studies (~70% accuracy) (62,63)

H2.2: Algorithms developed with free-living accelerometer data will classify activity type in free-living older adults more accurately than lab-based algorithms developed in Study 1

Exploratory Analysis:

Locomotion speed predicted by a machine learning algorithm will be correlated to speed in the 400m-walk from Study 1. The purpose of this analysis is to examine if locomotion speed predicted by machine learning algorithms may be used as a marker of physical function in free-living conditions. This will be the first study examining this association. If significant results are found, it may indicate that measuring free-living locomotion speed using machine learning algorithms is an alternative to assessing speed during a 400 m-walk.

CHAPTER II

LITERATURE REVIEW

Introduction

The number of people aged 65 and over accounted for 13% (39 million people) of the United States (US) population in 2008. It is estimated that in 2030, the number of older adults will represent 20% (~72 million people) of the US population (Figure 2.1) (7). This growing number of older people will create increased demand on the health care system, and, consequently, increase the economic burden on national health care.

Physical disability[2] is a special concern in late life. According to the US census, 28.6% of people aged 65 and over had a physical disability in the year 2000 (69). In addition, 9.5% of people in this age-range were unable to perform self-care tasks (69). Physical disability reduces quality of life and life expectancy in older people (10). Preventing physical disability and its related outcomes has grown as a public health concern over the last few decades (11). In 2000, the economic costs of long-term care for older adults with disabilities reached $123.1 billion dollars in the US. It is estimated that in 2040, these costs will equal $346 billion dollars (70). Preventive strategies will be more important than ever in reducing these costs and providing better health-related quality of life to older adults.

In this respect, physical activity is an effective mode of preventing and/or delaying the onset of physical disability. Older adults who are physically active have less chances of becoming physically disabled compared to their sedentary peers (11). Studies have shown that PA is associated with lower risk for physical disability and/or

[2] Functional limitation expressed in a social context, for example, inability to shop independently (4).

attenuation of the disablement process (11). Boyle et al. (71) reported that the risk of disability in activities of daily living (ADLs) were 7% lower for every additional hour of PA that older adults (80.5 ± 7.1 years old) performed per week. Another study (72) examined the influence of PA level on presence of disability prior to death in a subsample of the Established Populations for Epidemiologic Studies of the Elderly (EPESE). The most active older adults (men: 80+ years; women: 85+ years) were 2.43 times less likely to die with a disability than their sedentary counterparts. In a study by Miller et al. (73), older adults who walked one mile at least once a week were 9 to 36% less likely to become disabled compared to those walking less than a mile per week. These findings suggest that PA is effective in reducing risk for physical disability.

One of the mechanisms by which PA prevents physical disability is through the attenuation of age-related declines in the physiological systems. These declines are usually related to the presence of functional limitations[3] (10,11). It has been reported that participation in different activity types are associated with better preservation of the cardiorespiratory, musculoskeletal, and neuromuscular functions (74–79). Research has also shown that PA is associated with reduced risk for functional limitations, especially in the lower body (33,80). For example, Visser et al. (33) found that during a 4.5 year follow-up, inactive older men and women (70 to 79 years) were 1.47 and 1.44 times more likely to have incident mobility limitation (inability in completing a 400m-walk) compared to their active peers. Similarly, another study reported that older adults classified as inactive were 1.7 (men) and 2.1 (women) times more likely to suffer from lower extremity limitations compared to those classified as exercisers (80). In terms of

---

[3] Restrictions in performing fundamental physical and metal actions used in daily life by one's age-sex group (3).

meeting the United States physical activity recommendations, it was found that older adults who participated in more than 150 min/week of moderate-to-vigorous PA (MVPA) had better lower extremity function than those who participated in less than 150 min/week of MVPA (81).

Although this evidence indicates that PA is related to high levels of physical function, a more comprehensive understanding of the relationship between PA and physical disability risk has not been established in older adults. Available PA assessment methods, such as questionnaires and traditional accelerometer linear PA prediction models, have limited accuracy. This prevents accurate investigation of the dose-response relationship between PA and risk for physical disability (14,52,53,82).

Thus, it is essential to improve free-living PA measurement in older adults. It is particularly important to assess activity type in order to explore how daily performance of habitual activities may be related to the risk for physical disability. Recent advances in the objective PA assessment field revealed the potential of new methods in assessing activity type. For instance, studies using machine learning techniques were successful in identifying different activity types using artificial neural networks (ANN) and decision tree classifiers (25–27). In the study by Staudenmayer et al. (26), an ANN was also more accurate in predicting PAEE compared to three linear regression models. These results suggest that machine-leaning techniques can be used to provide a more comprehensive and accurate measure of free-living PA in older adults.

In addition, machine learning techniques can be used to improve assessment of sedentary behavior (SB), which is of particular importance as the modern society is becoming increasingly sedentary; especially the older segment of the population (19,83).

Accurate assessment of SB is important in order to further clarify its potential hazards to health. Sedentary behavior has already shown adverse effects on cardiovascular and metabolic health (84–87). It has also been associated with all-cause and cardiovascular mortality (88). Improving SB assessment will be necessary to proceed with the investigations of the aforementioned associations as well as to design prevention programs and public health policies more efficiently. In older adults, an accurate measurement method will also allow for better quantification of the influence of SB on physical function decline in older adults. In the next two sections, the limitations of common PA assessment methods and the current state of SB assessment are described (last section).

Limitations of Commonly-Used PA Assessment Methods

Physical Activity Questionnaires

Questionnaires have been extensively used to assess habitual PA behavior in epidemiological studies. These instruments are usually inexpensive, brief, and can provide a broad range of PA information (e.g. PA in different domains, PA during past year, PA-related energy expenditure, etc) (14).

Nevertheless, PA questionnaires have limited reliability and validity (14). Compelling evidence of the inaccuracy of PA questionnaires can be found in the systematic review by Prince et al. (82). The authors reported that questionnaires are poorly correlated ($r=0.37 \pm 0.25$) with objective PA measures such as doubly labeled water, heart rate, and accelerometers (82). The inaccuracy of questionnaires is largely due to their subjective nature. Respondents typically report PA based on individual perceptions and psychological factors (16). For the former, fitness level and previous experience influence how an individual perceives and reports PA intensity. For the latter, cognitive processes such as encoding, storage, retrieval and reconstruction of past information contribute to increased inaccuracy in self-reporting PA (16). In addition, another psychological factor that influences self-reporting of PA is social desirability. In this particular case, individuals tend to overestimate duration, intensity, and frequency of PA in an attempt to conform to 'socially acceptable norms' (14). All of these aspects can be more pronounced in older adults as they frequently suffer from mild cognitive impairments (17).

In addition to these factors, studies in the past have used age-neutral questionnaires to assess PA in older adults. These questionnaires are not appropriate for older adults because they lack questions regarding activities of daily living (ADLs), which are deemed crucial to assessing PA in this age group (18). Nonetheless, even questionnaires that were specifically designed for the older population do not result in substantial measurement improvement. In fact, their outputs were also poorly correlated with direct PA measures (r=0.11 to 0.32) (15,89). An explanation for this inaccuracy is that ADLs are recalled less accurately than exercise activities (14). According to Baranowski (16), it is easier to recall information about events that occur less frequently as opposed to those that occur more regularly. On a daily basis, ADLs and sedentary activities account for the majority of a person's day and are performed intermittently. This makes it difficult to accurately recall the time spent in each of those behaviors.

Due to their low accuracy, questionnaires have limited applicability in measuring PA changes resulting from intervention programs. In addition, when applied to studies investigating relationships between PA and health outcomes, large samples are necessary to achieve sufficient statistical power and minimize precision issues (14).

Given these limitations, it is difficult to obtain accurate PA measures in older adults when using questionnaires. Inaccurate PA measures prevent the investigation of how free-living PA relates to physical function in older adults. Therefore, other alternatives should be sought in order to more accurately measure free-living PA in older adults.

Pedometers

Pedometers are low-cost devices that count the number of accumulated steps during wear-time. They are usually attached to the waistband at the midline of the right or left thigh (90). Typically, pedometers use a spring-suspended lever system or an accelerometer-based internal mechanism to count steps (90–92). With the former system, vertical movement of the body leads to displacement of a lever that opens and closes an electrical circuit. Every time this cycle occurs, a step is registered. In contrast, accelerometer-based pedometers register steps in response to a given body acceleration threshold. These types of pedometers have an internal mechanism where a horizontal beam is attached to a piezoelectric crystal. In response to body acceleration, the horizontal beam bends and the piezoelectric crystal generates a voltage signal that is proportional to such bending (90–92). If the voltage signal is higher than a pre-determined threshold, a step is registered.

In research, both types of pedometers have been used to objectively assess PA in adults (91,93). Studies have shown that pedometers are more accurate than questionnaires in measuring PA (94,95). Despite higher accuracy, there are several limitations that prevent researchers from obtaining important measures of free-living PA when using pedometers. While these devices are adequate to measure walking behavior, they do not provide any output that can be used to assess activity type (90).

In older adults, pedometers can even be problematic in measuring walking behavior. This is because pedometers do not accurately capture slow gait and may also produce erroneous measures for shuffling patterns (96,97). In a study by Le Masurier et al. (98), a commonly used pedometer (Yamax SW-200) underestimated steps taken at 0.9 $m \cdot s^{-1}$ by 25% during a 5-min bout on a treadmill. Similarly, a study conducted by Storti et al. (97) found that, during a 100-step test, a Yamax digiwalker pedometer underestimated steps by 31.2% when gait speed was <0.8 $m \cdot s^{-1}$.

The inability of pedometers to detect slow walking speeds highlights the need for better instruments to assess walking behavior in older adults. Although frail older adults may move slowly and still keep their functional independence, it is known that frailty[4] is a pre-condition for loss of physical independence (60). Therefore, the assessment of free-living walking pattern is extremely important for a better understanding of its relationship with the physical disability process.

Accelerometer-Based Activity Monitors

Accelerometer-based PA monitors have emerged as a feasible option for objective assessment of free-living PA behavior, particularly due to their ability to provide more information about free-living PA behavior patterns (90,97,99). Currently, most of these monitors are lightweight devices that are usually worn on the hip and produce an output in response to acceleration resulting from body motion (24). Accelerometer-based PA

[4] Losses of physiologic reserve that leads to increased risk of disability (5).

monitors are classified according to the number of axes they detect acceleration from: uniaxial (single axis), biaxial (two-axes), or triaxial (three-axes). In terms of technical specifications, monitors may utilize different sensor mechanisms (e.g. piezoelectric, piezoresistive and capacitive sensors), different measurement ranges (g force magnitude) and sampling frequencies (100). In general, the acceleration detection and output production include three stages. First, a voltage charge proportional to the body acceleration is generated. An analog-to-digital converter then digitizes the signal and either a pre- (most commonly with commercial accelerometers) or post- filtering process is used to obtain the final output (100). This output, however, is not a physiological metric and needs to be processed by prediction models in order to be translated into measures of energy expenditure (EE) or thresholds of PA intensity (24,101).

To date, the most common method of translating accelerometer output into a PA metric is by using linear prediction models. These techniques, however, have important limitations (24). The next topic describes the process of developing simple linear prediction models and their main limitations.

Accelerometer Linear PA Prediction Models

These prediction models have been derived from accelerometer calibration studies carried out in laboratory settings (65,66,101,102). In these studies, participants usually perform different activities wearing hip-mounted activity monitor(s) while energy expenditure is measured by indirect calorimetry. Data are then used to develop linear regression equations that convert acceleration data into physical activity-related energy expenditure (PAEE) metrics (e.g. METs, $Kcal \cdot min^{-1}$). This approach has been useful in generating prediction models that can be applied to PA measurement in the field.

Nevertheless, the use of linear prediction models to process data from hip-mounted accelerometers generates inaccuracies in measuring intensity of household activities, which are usually intermittent and require limited movement of the hip (24,52,53). An example of this inaccuracy can be seen in Figure 2.2 where the *x-axis* of the figure shows activity counts (counts per minute) for different activities and the *y-axis* depicts measured METs for these activities. By applying the cut-point for light activity developed by Freedson and colleagues (101), a considerable number of ADLs that are of moderate intensity are misclassified as light intensity activity.

Therefore, employing linear prediction models to assess PA in older adults is problematic given that ADLs represent a substantial portion of their daily PA. In addition, linear prediction models do not provide any information about activity type, an important component of the interaction of older adults with their environment. As such, identifying activity type is essential for early identification of older adults at risk for physical disability. It is, thus, promising to employ advanced techniques to process accelerometer data in older adults.

Sedentary Behavior Measurement: Current Limitations and Future Directions

Sedentary behaviors (SB) are defined as activities with energy requirements of <1.5 MET such as sitting, lying down, and reclining (37). Sedentary behavior has also been measured with self-report methods and accelerometer cut-points (37,39,103,104). These two methods present limitations that restrict further investigations of the adverse effects of SB in older adults. Currently, the most common self-report SB methods include proxy measures of TV viewing time and sitting time (84,85,88,103). As previously mentioned, self-report methods are problematic in older adults, especially in those with cognitive impairments (16,17). Aside from this fact, proxy measures may lead to substantial misclassification of light activities and sedentary activities due to misjudgment from the respondents. This occurs because people may not have a clear discernment of the difference between sedentary and light activities, or because they may find it difficult to recall temporal information about unstructured activities (16).

In contrast, accelerometer SB cut-points may substantially overestimate or underestimate time spent in sedentary activities due to systematic misclassification, which arises from the lack of specificity of the cut-points (39). This misclassification is again more common between sedentary activities and lifestyle activities (24), which are the two groups of activities most often performed by older adults (12). Classifying sedentary activities as lifestyle activities or vice-versa can affect the investigation of the associations between SB with adverse health-related variables.

Thus, it is important to accurately measure time that older adults spend in SB as well as specific characteristics of these behaviors. These characteristics can be used to design interventions that are more effective at increasing PA and to investigate the adverse effects of SB on physical function. Therefore, there is great potential to further explore the use of objective methods to assess SB in older adults in the coming years. The use of accelerometer-based PA monitors and advanced statistical methods to process acceleration signals provides an appropriate combination to move the field forward.

Machine learning Methods: An Alternative to Improve Free-Living PA and SB Assessment in Older Adults

Machine learning methods can be defined as computational adaptive methods that are able to automatically improve performance when provided with examples (training data). They are appropriate for solving non-linear functions, especially when the data set demonstrates patterns that are complex (32). The main objective of using machine learning methods is to create computer algorithms that can classify outputs based on input vectors. These algorithms can be developed and trained using different types of learning concepts, but the two most common are *supervised learning* and *unsupervised learning* (32). In supervised learning, algorithms are trained with data that has examples of input vectors with their corresponding output vectors, or in other words, the data set is labeled. In unsupervised learning, the training data is comprised of input vectors with no labeled outputs. Algorithms developed via unsupervised learning have the goal of discovering groups of similar examples, usually clustering them according to their proximity in the input space (105,106). In activity classification, it is more common to use supervised learning versus unsupervised learning algorithms. Therefore, this section will only describe the use of the former in classifying activity from acceleration signals.

In a comprehensive review, Preece et al. (32) listed various studies that successfully used different machine learning methods (decision tree classifiers, ANNs, SVMs, etc.) to process accelerometer data. These methods have potential to improve PA and SB estimates in older adults, especially by predicting activity type. This information is essential to understanding the role of different PA and SB modes on health outcomes and physical disability risk in older adults. A brief overview of the development process of a machine learning-based activity classification algorithm is provided below.

First, the appropriate windowing technique for the sensor signal is selected; this can be a fixed window (e.g. min by min), event-based window (e.g. toe off, heel strike), or activity-defined window (e.g. bout of activity). Subsequently, different types of signal features (e.g. time- and frequency-domain features, and time-frequency features) are extracted according to the selected windowing technique. Statistical analyses are then used to identify the best candidates to be used as input features for the activity classification algorithm. In general, optimal features are those with high intra-class and low inter-class correlations. The input features are then used as predictor variables to develop and train the activity classification algorithm. Once developed and initially trained, accuracy of the algorithm can be improved with further training (data input). Therefore, as more data are inputted into the model, its ability in discerning between signals corresponding to different activities is optimized (32).

Farehnberg et al. (107) were early adopters of a machine learning method to identify different postures and motions from acceleration signals. In a series of studies (62,107,108), they used stepwise discriminant analyses to process data from a multichannel piezoresistive accelerometer to identify eight to nine different postures/activities. Since then, many researchers have made use of machine learning methods to develop classification algorithms for postures and ambulatory motion (109). More recently, advances in computational power have led to the development of more sophisticated classification algorithms, thus, allowing researchers to identify a wider variety of activities using acceleration signals (110).

In a study by Bao and Intille (25), a decision tree classifier was used to classify activity type based on different signal features (e.g. mean, energy, frequency-domain entropy) from five accelerometers placed on different parts of the body. The classifier was able to identify 20 activities with an accuracy rate of 84%. An important aspect of this study was the ability of identifying diverse activities such as sedentary, household, locomotion, leisure-time and exercise activities. In another study, Tapia et al. (40) developed and trained a fast decision tree classifier for real-time recognition of 30 gymnasium activities (e.g. rowing, bicep curls, push-ups, walking, etc). Data were collected on 21 participants who wore five triaxial accelerometers on different parts of the body and a heart rate monitor. The authors used different time- and frequency-domain features for the algorithm and the activity classification was done in 4.2 s windows. The classifier had an 80.6% overall accuracy in recognizing activity type.

Despite these promising results, few studies have used machine learning methods to process the output of commonly used accelerometers. Pober et al. (111) employed quadratic discriminant analysis (QDA) and hidden Markov Modeling (HMM) to classify activity type using data from the ActiGraph 7164 activity monitor. The overall classification accuracy of QDA and HMM in classifying four different activity types (walking, walking uphill, vacuuming and computer work) were 70.9 and 80.8%, respectively. In 2009, Staudenmayer et al. (26) developed an ANN to predict activity type and physical activity-related energy expenditure (PAEE) using data from the ActiGraph 7164. The recognition rate of the ANN in predicting activity type was 88.8% (95% CI: 86.4 - 91.2%), with most activities (11 out of 18) being correctly identified more than 90% of the time. In terms of PAEE prediction, the performance of the ANN was promising, with small prediction errors. The ANN measurement bias for PAEE was no greater than 0.10 METs and the largest root mean squared error (RMSE) was only 1.22 MET. When three different linear regression methods were used to predict PAEE from the same accelerometer data, the measurement bias was as large as 1.21 MET and the largest RMSE was 2.09 METs. More recently, a study by De Vries et al. (27) developed an ANN to predict activity type using data from an ActiGraph GT1M worn on the hip and another on the ankle. The overall performance of the ANN in recognizing activity type was 83.0%. An important result from both studies was the success in identifying SB. While Staudenmayer et al. (26) reported 88% correct classification rate for sedentary and light activities, De Vries et al. (27) reported that 'sitting' was correctly classified 90.6% of the time.

The results achieved by the different studies indicate that machine learning methods can be used to improve free-living PA and SB estimations in older adults. A recent meta-analysis on the potential of using accelerometry to assess activity type in older adults reviewed several studies using activity classification algorithms (109). Few studies were conducted in samples of older adults and they were mainly concerned with static postures and ambulatory activity. For example, Culhane et al. (112) developed a threshold-based algorithm to discern between lying, sitting, standing, and dynamic motion. The algorithm was able to identify these activities with a 92% correct classification rate in a rehabilitation setting. In another study, a threshold-based algorithm was employed by Bourke et al. (45) for fall detection in older adults. The authors were able to attain an accuracy rate of 100% in differentiating fall events from ADLs using measures from a trunk-mounted accelerometer in older adults. Using a Wavelet

Transform algorithm, Najafi et al. (44) were successful in identifying postures (sitting, lying, and standing), transitions (sit to stand, and stand to sit), and walking in older people. Specificity rates in the free-living environment were 92.1, 93.4 and 99.7% for sitting, standing+walking, and lying, respectively.

Although assessment of postures and ambulatory motion is important, it is necessary to develop methods that can identify a broader range of activities in order to allow a more comprehensive understanding of the associations of PA and SB with physical disability risk and health outcomes in older adults. Another relevant aspect that becomes clear from the meta-analysis is the need to develop activity classification algorithms for commercially available PA monitors. This is crucial if PA and SB measurement are to be conducted in large-scale studies. The studies by Pober et al. (111), Staudenmayer et al. (26), and De Vries et al. (27) demonstrated that placement of one or two commercially available PA monitors was sufficient to produce accurate predictions of activity type.

Considering the current evidence, it is necessary to expand the use of machine learning methods to develop accelerometer-based activity classification algorithms in older adults. In addition, utilizing commercially available accelerometer-based PA monitors will enhance feasibility of these algorithms in free-living older adults. The immediate applications of activity classification algorithms are innumerous in older adults. There will be special implications for understanding the relationship of free-living PA and SB with physical function. Some of these applications are discussed next.

Applications of Accelerometer-Based Activity Classification Algorithms

Free-living Physical Function Assessment

Currently, physical function is usually assessed using self-report instruments (46,49) and physical performance tests (47,48). Although these methods are useful in screening people at risk for physical disability, they do not provide direct information about how an individual interacts with his or her environment. Accelerometer-based activity classification algorithms can be partially used to obtain such information. They may be used to quantify engagement of older adults in different activity types in free-living conditions.

By using activity classification algorithms, it is possible to identify activity characteristics that are currently only measured in constrained tasks. For example, walking speed, which is an important predictor of mortality and physical disability, is usually measured with a 400 m walk test (113,114). Machine learning techniques can be used to develop activity classification algorithms for identifying ‘free-living’ walking speed from acceleration signals. Studies have been successful in identifying various walking and running speeds employing such techniques (40,56). In addition, machine learning techniques can be used to create specific algorithms to identify patterns of ambulatory activity (e.g. time series variability). This information is important considering that older adults who are physically active present more complex patterns of ambulatory activity than those of inactive older adults (115). Therefore, identifying ambulatory patterns provides an additional form of screening those at risk for physical disability.

Another functional task that is typically assessed with a physical performance test is the ability to stand from a seated position (47,48). In this case, the time that a person takes to complete a particular number of chair rises (e.g. five) is scored. This score is an important indication of lower extremity function (47). However, the score only provides a way of assessing physical impairment[5]. Using objective monitoring to assess how frequently an individual executes sit-to-stand transitions in free-living conditions may help to better understand how physical environmental demands play a role in maintaining good physical function. Activity classification algorithms have the potential for assessing such transitions when properly trained (44).

Similarly, sedentary behaviors such as lying, reclining and sitting can be identified with machine learning methods. Studies have developed activity classification algorithms that were accurate in identifying activities such as lying, sitting, watching TV, computer work, and reading (25,40,44,62,116,117). This demonstrates the potential of machine learning techniques to improve upon over simple cut-points in assessing SB in free-living conditions.

Finally, a limitation of previous PA assessment methods for older adults was their inability to accurately assess low intensity PA (e.g. lifestyle activities). Machine learning techniques can identify these activities from acceleration signals. The studies by Staudenmayer et al. (26) and Bao and Intille (25) demonstrated that several activities could be correctly identified using either an ANN or a decision tree classifier. To date, ability to perform ADLs is usually assessed with self-report instruments such as the SF-36 or battery tests encompassing ADLs such as the ‘Continous-Scale Physical Functional

[5] Abnormalities at the level of tissues, organs, and body systems (6).

Performance test' (46,49). Ability in identifying ADLs may be important to identify older adults at immediate risk of becoming physically dependent. Low engagement in lifestyle activities with concomitant increase in sedentary behavior may indicate higher risk for disability in ADLs. This type of disability is a close indicator of the risk for severe disability, and, consequently, for physical dependence (61). Thus, the development of activity classification algorithms will aid in assessing risk of ADL disability.

Dose-response Relationship between Free-Living PA and Physical Function

With aging, several changes take place in the human body and lead to considerable loss of physiological function in late life. For example, in comparison to middle-aged adults (46 years old), older adults (78 years old) experience reductions of approximately 40-45% in muscle strength of knee extensors and flexors (118). Similarly, changes in the cardiorespiratory system can result in declines as large as 40% in maximal oxygen consumption (VO_2 max) from age 25 to 65 (119). A less steep decline is observed for bone loss, which occurs at a rate of approximately 0.5% per year after the age of 40 (76). Nevertheless, this bone loss is sufficient to increase the chances of bone fractures in older adults, especially in women (76).

Due to these changes, older adults are usually at increased risk for functional limitations. Physical activity is an important factor in minimizing this risk. Thus, the ability in measuring activity type is important to quantify particular PA episodes that create specific stress on the different physiological systems (e.g. musculo-skeletal, neuromuscular and cardiorespiratory systems). In this respect, studies using accelerometer-based activity classification algorithms have been able to identify diverse aerobic activities (e.g. walking, running, biking) and neuromuscular activities (e.g.

strength-training, sit-to-stand transitions, push-ups) in younger adults (25–27,40). Developing similar methods for older adults will allow researchers to establish a dose-response relationship between free-living activity type and physical function. This information is vital to making more appropriate PA recommendations for preventing functional limitations, and, consequently, physical disability in older adults.

Assessment of Free-living PA and SB in Large-Scale Studies

Obtaining accurate measures of free-living PA and SB is imperative before implementing any public health policies for PA promotion in older adults. For instance, it allows for setting realistic PA recommendations that can be achieved by most of older adults. In addition, it allows for detecting epidemiological PA trends arising from real changes rather than artifact in the data due to inaccurate measures. With the advancement of feasible body-worn sensors, it became possible to objectively assess PA and SB in large-scale studies. In 2003, the *'National Health and Nutrition Examination Survey'* (NHANES) used an accelerometer-based PA monitor to objectively assess PA and SB in a nationally representative sample of Americans (19,37). However, simple cut-points were applied to process accelerometer data from that study (19,37). As mentioned before, the cut-point method is unable to accurately detect lifestyle activities due to the non-linearity between accelerometer hip data and PAEE (24,52). The cut-point method may be especially inaccurate in older adults as they spend substantial time performing lifestyle activities.

It is likely that data from NHANES suggesting that older adults spend 8.6 $min\cdot day^{-1}$ in moderate PA is incorrect (19). Studies have shown that several lifestyle activities of moderate intensity (e.g. sweeping, window cleaning, gardening, lawn mowing, raking, etc) produce accelerometer outputs that are typically classified as light intensity PA by the “cut-point” technique (120,121). In contrast, static upright posture as well as some other lifestyle activities that demand little hip movement (e.g. washing dishes, folding clothes) may be misclassified as sitting (24). Therefore, in the study by Matthews et al. (37) significant bias may have resulted from using a simple cut-point (<100 counts per minute) to process accelerometer data from a hip-mounted activity monitor. This indicates the importance of employing advanced techniques to process accelerometer data in order to obtain accurate free-living PA and SB measures. Machine learning methods have been accurate in predicting both activity type and intensity (26). The ability to assess activity type from acceleration signals is an advantage of using machine learning techniques and may be applied to obtain a more comprehensive measure of free-living PA and SB behavior of older adults. This is important before making public health policies that specifically target ways of increasing PA level in such population.

Low Intensity Physical Activity and Health Outcomes

Accurate PA measures in older adults are also important to establish a dose-response relationship between low intensity PA and health outcomes. Higher levels of PA have been related to positive health outcomes such as prevention of diabetes and cardiovascular disease (30). However, it is possible that such relationships have been underestimated given that previous methods were unable to accurately capture low intensity PA, such as lifestyle activities.

The importance of low intensity PA had gained attention over the last few years when studies found that SB has major adverse effects on metabolic health (84–87). In this sense, low intensity PA is a feasible strategy of reducing SB, and, consequently, of avoiding its negative effects on health. This concept became stronger with a study finding that even small breaks in sedentary time are related to better metabolic profile (38). Machine learning methods may provide accurate measures of low intensity activities such as lifestyle activities. Using these methods will allow for reinvestigating the relationship between low intensity PA and metabolic health in older adults. This will have important implications for the development of future recommendations on avoiding sedentary behavior in older adults.

Figures

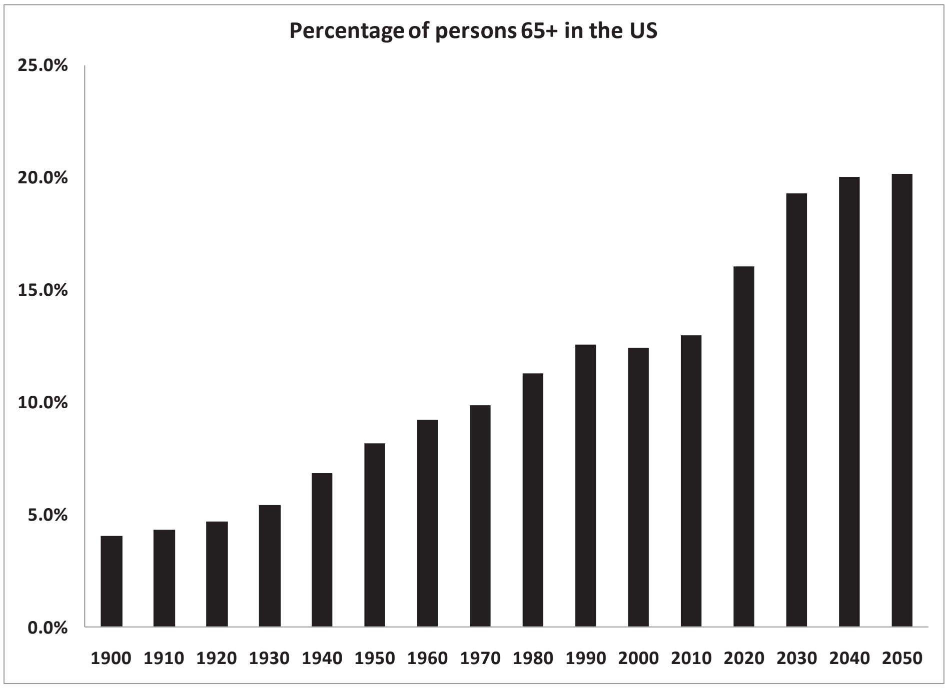


Figure 2.1: Percentage of persons aged 65 and over in the United States from 1900 to 2050. Source: US Census Bureau, 2003 (69).

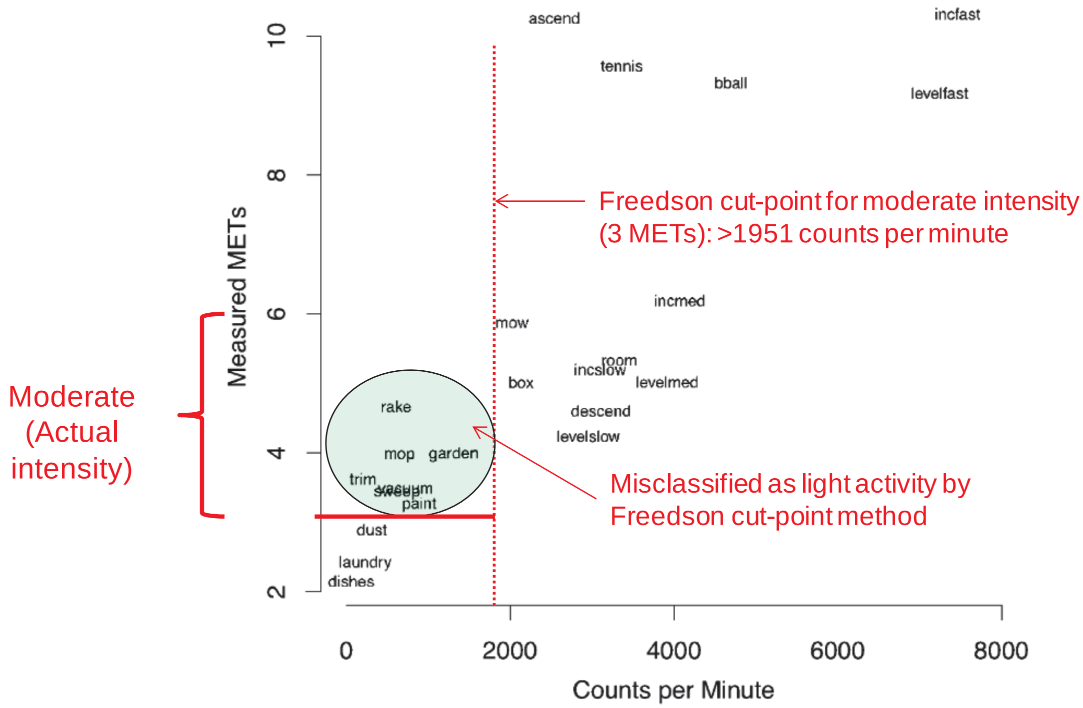


Figure 2.2: Misclassification of intensity of ADLs by a commonly used accelerometer cut-point method. Source: Adapted from Kozey et al. 2010 (120).

CHAPTER III

METHODS

<u>Study 1: Development and Validation of Accelerometer-Based Activity Classification Algorithms for Older Adults</u>

Recruitment, Eligibility and Screening

Forty healthy and ambulatory older adults (20 females and 20 males) in the age range of > 65 to 80 years will be recruited to take part in this study. Volunteers will be recruited from Amherst and surrounding areas using flyers, short articles on different media outlets (e.g. University website, local news), visits to senior centers, and word of mouth. Volunteers will be screened over the phone and will be automatically excluded if they present with any of the following conditions: 1) congenital heart disease 2) myocardial infarction or stroke in the past year, 3) congestive heart failure, 4) chronic obstructive pulmonary disease, 5) insulin-dependent diabetes mellitus, 6) Parkinson's disease, 7) Alzheimer's disease or any type of dementia, 8) active cancer treatment (e.g. radiotherapy, chemotherapy), 9) Liver and/or kidney disease, 10) Epilepsy, 11) current use of 5 or more prescribed medications that affect metabolism or cardiovascular and hemodynamic responses to exercise, and 12) use of any ambulatory assistive device. If volunteers are considered eligible, they will be invited to the *Physical Activity and Health Laboratory* for an informed consent visit.

During the informed consent visit, a researcher will explain the study and answer any questions the volunteers have. They will be informed that in order to be completely eligible for the study, they will need to complete a short physical performance battery test and obtain medical clearance after signing the informed consent document. If they decide to participate in the study, they will sign the informed consent document and complete questionnaires about their personal health history, physical activity readiness, physical activity status (NASA physical activity scale), habitual physical activity and physical function level (SF-36) (Appendices B-G). Volunteers will then complete the *Short Physical Performance Battery Test* (47). The test is composed of the following 3 activities: a) balance test - ability to stand with the feet together in the side-by-side, semi-tandem, and tandem positions, b) time to walk 8 feet, and c) time to rise from a chair and return to the seated position 5 consecutive times. For each activity, participants will receive a score of 1 to 4 based on their performance compared to normative values (47). The scores on the three activities will be summed to produce the final performance score (maximum score 12). All volunteers will be required to score 12 in order to be considered for the study. Volunteers scoring less than 12 will be excluded from the study for minimizing chances of enrolling participants with mobility-impairments. Once the informed consent visit is completed, we will request volunteers to obtain medical clearance from their physician in order to proceed with participation in the study. A researcher will explain the *Medical clearance form* (see Appendix H) as well as the reasons for obtaining it. If they agree, a request form will be faxed to their physician who can approve or disapprove their further participation in the study. If granted approval, participants will be scheduled for the *Activity routine visit*, which will be conducted in the *Physical Activity and Health Laboratory*.

Activity Routine Visit

Resting Metabolic Rate

Participants will refrain from consuming any food, beverages (other than water), and caffeine for 4 h prior to the visit. In addition, they will be asked not to exercise on the same day of the visit. Once they report to the laboratory, participants will sit and remain quiet for a 5-min period before heart rate and blood pressure are measured. They will proceed with the visit if the following criteria are met: 1) heart rate below 100bpm, 2) systolic blood pressure below 140 mmHg, and 3) diastolic blood pressure below 90 mmHg. Participants will rest in a supine position for at least 15 min before resting metabolic rate (RMR) is measured. A MedGem Analyzer (Healthe Tech, Inc, Golden, CO), which is a handheld portable indirect calorimetry system, will be used to measure resting metabolic rate. Before each measurement, participant information will be inputted into proprietary software and a disposable mouthpiece will be attached to the handheld device. The MedGem will then be positioned on a solid surface to be calibrated and initialized according to the manufacturer specifications. Once calibrated and initialized, the handheld device will be given to the participants who will be asked to breathe normally through the mouthpiece for a period of 10-15 minutes. Validity and reliability of the MedGem Analyzer have been demonstrated in adults in previous studies (122).

Instrumentation

Following RMR measurement, participants will be fitted with three ActiGraph GT3X+ activity monitors (ActiGraph, Pensacola, FL). This device is a lightweight triaxial PA monitor (4.6cm x 3.3cm x 1.5cm, 19g) that measures acceleration ranging in magnitude from -6 to +6 g's. The accelerometer output is sampled at 30 to 100Hz and digitized by a twelve-bit analog-to-digital convertor. All GT3X+ monitors will be synced to the same laptop and initialized in advance to collect data at a sampling rate of 80Hz. They will be positioned on the dominant wrist, ankle and hip of the participants.

In addition, participants will wear the Oxycon Mobile indirect calorimetry system (Carefusion, Yorba Linda, CA). This system collects breath-by-breath data and requires participants to wear a facemask and two small units mounted on a harness secured to the upper back. The flowmeter and gas analyzer units of the Oxycon Mobile will be calibrated using a 3-liter air syringe and a known gas mixture (16.03% O_2 and 4.02% CO_2). Validity and reliability of this instrument in measuring oxygen consumption of adults over different exercise intensities has been demonstrated in the study by Rosdahl et al. (123).

Procedures

First, participants will perform three postures in the following order: sitting still, standing, and lying down. Each posture will be performed for 30 seconds with no interval in between. Participants will then be assigned to perform one of two activity routines (Table 3.1). Each activity will be performed for 5 min and a 4-min rest will be allowed after completion of every activity.

Rate of perceived exertion (RPE) for the participants will be assessed using the Borg scale. Assessments will occur after each activity. The scale contains numbers from 6 to 20 that correspond to different levels of exertion (Appendix J). Before starting the activity routine, we will instruct participants on how to rate their exertion level on the RPE scale (Appendix J). The Borg scale has been shown to be valid and reliable in older women aged 75.5 ± 3.8 years (124).

Data Processing and Statistical Evaluation

Raw acceleration signals (g) from the three activity monitors (i.e., hip, wrist, and ankle) will be synchronized to the corresponding activities. Signals will be labeled according to the individual activity type and activity groups (e.g., Sedentary, Household, Locomotion). Once data are properly cleaned and labeled, a visual inspection will be carried out in order to ensure data are properly aligned. Time-domain features and frequency-domain features for these acceleration signals will be extracted for every 20-second window. For obtaining steady-state metabolic data, the first two minutes of data for each activity will be discarded. The VO_2 values of the remaining 3 minutes will be averaged and then divided by 3.5 in order to calculate METs for each activity. Acceleration features along with metabolic data from the different activities will be inputted into three different machine learning models, namely Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF) models. Activity classification algorithms for each type of model will be developed using data from individual monitors (e.g., hip alone, wrist alone) and combined monitors (e.g., hip and wrist, wrist and ankle). We will develop algorithms using only time-domain features and also using both time- and frequency- domain features. For prediction of activity intensity (METs and multiples of RMR), the regression versions of the models will be used (e.g., Support vector regression, random forest regression). METs for each activity will be calculated by dividing steady-state activity VO_2 (minutes three to five) by 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$. For calculating multiples of RMR ($Mult_{RMR}$), we will divide steady-state activity VO_2 by participant's resting VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$). Time- and frequency- domain features of the acceleration signals along with MET values or $Mult_{RMR}$ will be input features for developing algorithms to predict METs or $Mult_{RMR}$.

We will also develop models to estimate locomotion speed. Speed during the 400 m walk will be calculated for each participant using the following equation: Speed ($m \cdot s^{-1}$) = (400/time to complete test in s). Time- and frequency- domain features will be extract for acceleration signals from slow treadmill walking (0.8 $m \cdot s^{-1}$) and 400 m walk. These features along with corresponding locomotion speed (0.8 $m \cdot s^{-1}$ and speed during 400m walk) will be used to train the models for estimation of locomotion speed.

Statistical Evaluation

Performance of the algorithms for classification of activity type and prediction of activity intensity and locomotion speed will be determined using a ‘leave-one-out’ validation approach. The accuracy of activity type classification algorithms will be tested by calculating percent correct classification for activity group category and individual activity type. A confusion matrix to determine misclassified minutes across the different activity group categories will be computed for the algorithm with the best overall accuracy. Sensitivity and specificity for classifying the different activity group categories will be calculated for the algorithm with the best overall accuracy. Linear mixed models will be used to calculate measurement bias for time spent in different activity group categories as well as for prediction of METs and multiples of RMR. Significance will be determined by 95% confidence intervals. Coefficient of determination will be calculated for algorithms developed for prediction of locomotion speed.

Tables

Table 3.1: Activity routines

Routine 1	Routine 2
Crosswords	Playing cards
Self-care (miscellaneous)	Laundry
Organizing the room	Dusting
Gardening	Vacuuming
Carrying groceries	Slow walk (~1.8 mph)
400m walk	400m walk
Tai-Chi	Playing Bowling

Study 2: Validation of the Accelerometer-Based Activity Classification Algorithms in Free-living Older Adults

Recruitment

Twenty older adults (10 males and 10 females) who participated in study 1 will be invited to take part in study 2. A researcher will explain the study and answer any questions they may have. If they demonstrate interest in participating in study 2, the researcher will provide further information and answer any concerns. They will be given an informed consent document approved by the Institutional Review Board from the University of Massachusetts Amherst and sign and date it if they choose to participate in the study (Appendix K). The researcher will then schedule a 3h time block to directly observe the participants.

Instrumentation

Personal Digital Assistant

A Personal Digital Assistant (PDA) programmed for *continuous focal sampling DO* (CFS-DO) (*The Observer*®; Noldus Information Technology, Wageningen, The Netherlands) will be used to code the activities performed by the participants in the free-living environment. Three activity characteristics will be captured using the PDA and CFS-DO software:

1) Activity type - A menu of activities for the PDA will be created by two experts in the field before starting this part of the project and will contain activities that are commonly performed by older adults. The selection of appropriate activities to be included in the menu will be based on a literature search on time use at older ages (12).

2) Intensity range - four intensity categories will be available on the PDA: sedentary (<1.5 METs), light (1.5 to <3.0 METs), moderate (3.0 to <6.0 METs), and vigorous (≥6.0 METs). Activity intensity from activity type based on the Compendium of Physical Activity (1).
3) Activity duration: A 1-sec record interval will be used to record the activities.

Activity Monitors

Three ActiGraph GT3X+ monitors will be synced to the same computer and initialized in advance. The monitors will be initialized using the ActiLife 5 software to collect triaxial accelerometer data at 80 Hz. Before leaving the laboratory to meet with the participants, the observers will place the monitors onto an elastic belt (hip unit) and two elastic straps (wrist and ankle).

Heart Rate Monitor

A heart rate belt and a heart rate monitor RS400 (Polar Electro, Oulu, Finland) will be used as the criterion measure for activity intensity in the free-living setting. The heart rate monitor will be synced to the clock of the computer used to initialize the activity monitors and PDA.

Observers

Three observers will be trained on how to use the PDA and the continuous focal sampling software. They will receive instructions on how to code activity type and intensity during face-to-face training sessions and group discussion meetings. At the end of the training period, the observers will complete a test to examine inter-observer reliability. They will watch a video containing twenty activity clips in two occasions and code activity type and intensity for each clip in each occasion. Inter-observer reliability will be calculated by the kappa agreement test. A kappa coefficient of 0.8 or higher will be required before starting the study.

Direct Observation

The observers will bring the monitors and meet the participants at the pre-determined time and location. Before starting the DO session, observers will assist participants with placement of the monitors (dominant wrist, hip and ankle) and will make sure that they feel comfortable to proceed. Participants will be instructed to perform their daily routine as if no one is observing them. Once the participants are ready, observers will start the DO session and will record the activities performed by the participants during the 3h time block. With the CFS-DO software, observers will be able to record the activity type and activity intensity as they occur. Observers will also carry a memo-notebook and a pen to record any activities not listed on the PDA as well as to take notes about corrections to be made during the data entering process.

Data Processing and Statistical Analysis

Direct observation data will be downloaded to a laptop using the *The Observer*® software (*The Observer*®; Noldus Information Technology, Wageningen, The Netherlands). Text files containing the activities performed in the free-living environment will be generated for each participant. In each text file, the activities performed by the participant will have a time stamp and an intensity code. The activities will be collapsed into groups according to their type (e.g. sedentary, lifestyle, ambulatory, postural transitions) and intensity category (e.g. sedentary, light, moderate, vigorous). Total time spent in each activity type and intensity category will be quantified for each participant.

Accelerometer data will be downloaded to a laptop using the *ActiLife* software (ActiGraph, Pensacola, FL) and will be later extracted to match the corresponding DO time blocks. These data will then be processed using the activity classification algorithms developed in study 1 to derive total time spent in each activity type and intensity for each participant.

Activity type predicted by the activity classification algorithm will be compared to the DO data. Percent correct classifications will be calculated in order to assess the accuracy of the activity classification algorithms, both in terms of overall activity classification and activity group classification. In addition, bias of the activity classification algorithms in classifying time spent in each activity group type will be calculated.

CHAPTER IV

DEVELOPMENT AND VALIDATION OF ACCELEROMETER-BASED ACTIVITY CLASSIFICATION ALGORITHMS FOR OLDER ADULTS

Abstract

Purpose: To develop activity classification algorithms to process accelerometer data in older adults. **Methods:** Thirty-five healthy older adults (21 women and 14 men, mean ± SD age = 70.8 ± 4.9 years) wore a portable metabolic system to measure energy expenditure and three ActiGraph GT3X + activity monitors (dominant wrist, hip and ankle) initialized to collect data at 80hz. Participants performed sedentary (SED), locomotion (LOC), household (HOU), and recreational (REC) activities. Time- and frequency- domain features for each activity were extracted from the accelerometer signals of each monitor and steady-state METs were calculated from the portable metabolic system. These data were used to train artificial neural network (ANN), random forest (RF), and support vector machine (SVM) models for prediction of activity type and activity intensity. A leave-one-out method was used to test the accuracy of each model. **Results:** Accuracy of the models in detecting activity type ranged from 87% (ANN, RF, and SVM hip) to 96% (SVM wrist) using single monitor data. There was no substantial improvement in accuracy when combining data from two or three monitors (+ ~2%). The highest classification accuracy was for the SVM wrist algorithm (SED, LOC, HOU and REC activities classification accuracy: 97%, 97%, 96%, and 94%). The biases for MET prediction were small ranging from -0.01 MET (RMSE: 0.54 MET) for the RF wrist algorithm to 0.02 MET (RMSE: 0.67 MET) for the ANN hip algorithm. **Conclusion:** The activity classification algorithms in this study accurately predicted activity type and

intensity from a single accelerometer. Machine learning models for processing accelerometer data may be valuable tools for estimating METs and detecting activity type in free-living older adults.

Introduction

Exercise interventions typically employ structured exercise regimens to maintain and improve function of the cardiovascular, neuromuscular, and skeletomuscular systems in older adults (33–35,125–127). During these types of interventions, it is relatively easy to quantify frequency, duration and load of exercise. Engagement in structured exercise usually results in many health benefits but this type of activity is only performed during a small portion of an individual’s day. As a consequence, it is also important to assess physical activity (PA) occurring outside structured exercise. To date, self-report tools have been the method of choice to assess free-living PA behavior in older adults (11). Measures obtained with these tools have been used to examine the association of PA with risks for physical disability (10,11). It is imperative to improve PA assessment in older adults to obtain a more complete understanding of this relationship. Accelerometer-based activity monitors are ideal tools for responding to this need. However, the majority of research using commercially available activity monitors has employed linear regression methods to predict activity intensity from accelerometer data (19,21). Physical activity intensity has been the metric of choice because of the vast amount of research examining the association of PA intensity with health outcomes (30). Unfortunately, linear regression methods are especially inaccurate for activities of daily living and may lead o misclassifications of PA intensity in older adults (26,53).

More recently, the advent of more sophisticated activity monitors has allowed researchers to apply advanced statistical and computational methods to classify activity type from acceleration signals (32,41,110). The possibility of using these techniques in older adults is of interest to improve activity behavior assessment in this age group.

Previous studies have used machine learning techniques to process accelerometer data in younger adults. These studies were able to predict activity type with recognition rates higher than 80% and activity intensity with bias (average difference between predicted minus actual) as low as 0.1 MET (26,27,41,128). The machine learning techniques most frequently used to process accelerometer data have been artificial neural networks, support vector machines and decision tree classifiers (26,41,110) . A common characteristic of these techniques is the identification of complex patterns contained within acceleration signals for different activities (32). Algorithms developed using machine learning techniques improve performance with additional training data. The flexibility of these techniques is advantageous for processing large volumes of data, such as those generated by sophisticated activity monitors (e.g. ActiGraph GT3X+) that can collect raw acceleration signals at sampling rates of 100+ Hz. Recent studies have used these raw acceleration signals to classify activity type in younger adults (41,110).

Despite these major advances in younger adults, little progress has been made in using machine learning techniques to classify activity type and intensity from raw acceleration data in older adults. Currently, the studies predicting activity type in older adults have classified postures and gait using prototypes or accelerometers developed for clinical settings (109). Applying machine learning techniques to data from commercially available activity monitors may have important measurement implications for older adults. One example would be the use of activity type classification algorithms to estimate locomotion time and speed from accelerometer data. Both locomotion time and speed have been associated with physical disability, survival time and mortality in older adults (54,73). Therefore, objective detection of critical levels of locomotion time and speed could produce information for developing public health interventions that would benefit a large number of older adults.

Compliance with wearing activity monitors in free-living conditions may be influenced by monitor placement. In this regard, the NHANES sought to increase participants' compliance by adopting wrist as the placement site for activity monitors in their physical activity measurement study protocol (59). In addition, recent studies in younger adults developed algorithms to classify activity type from wrist- and ankle- worn activity monitors (41,59). The results suggest that placement of activity monitor on the wrist produce accurate classification of activity type whereas ankle placement results in lower but still acceptable activity recognition rates (41,59). In contrast, limited information is available on best monitor placement for activity type classification in older adults.

In view of the gaps identified in the literature, the purposes of this study were: 1) to develop and evaluate machine learning algorithms to predict activity type from wrist, hip, and ankle accelerometer data collected using the ActiGraph GT3X+ activity monitor in older adults, 2) to develop and evaluate machine learning algorithms to predict activity intensity from wrist, hip, and ankle accelerometer data, 3) to determine best monitor placement for activity type and intensity prediction from accelerometer data in older adults, and 4) to develop machine learning algorithms to estimate locomotion speed from accelerometer data in older adults.

Hypotheses

Hypothesis 1: Our machine learning models would accurately predict activity type in older adults (≥ 80% accuracy)

Hypothesis 2: Our machine learning models would predict activity intensity with similar accuracy as observed in previous studies using machine learning models in younger adults (bias < ±0.1 MET)

Exploratory Analyses

1) Studies in younger adults suggest that placement of activity monitor on the wrist or ankle produce accurate or acceptable recognition for activity type (41,59). There is limited evidence as to what monitor placement produces the best recognition rate for activity type in older adults. Monitor placement is important for increasing compliance and also for reducing burden on the participant. Given this importance, we determined optimal location for activity monitor placement based on the results from this study. It could not be hypothesized which placement would be superior for prediction of activity type and intensity since no similar studies have been conducted in older adults.

2) To date, few studies have estimated locomotion speed from acceleration signals and they were conducted in younger individuals or using accelerometers designed for clinical settings (56–58). Objective assessment of locomotion speed in older adults is of major importance, as studies have reported associations of locomotion speed with survival time, risks for disability, and risks for becoming frail (54,60,61). Thus, we developed machine learning algorithms for prediction of locomotion speed from raw acceleration signals in older adults from this study. This was treated as an exploratory aim since no measurement error values could be found in the literature to serve as reference values for the present investigation. We could not hypothesize the degree of accuracy and measurement error to be expected from the models in predicting locomotion speed.

Methods

Recruitment, Eligibility and Screening

Thirty-five healthy older adults were recruited from Amherst, MA and surrounding areas. Participants were recruited using flyers, short articles through different media outlets (e.g. University website, local news), visits to senior centers, and word of mouth. Exclusion criteria for this study included: 1) age <65 or >85 years, 2) diagnosed heart disease, 3) myocardial infarction or stroke in the past year, 4) congestive heart failure, 5) chronic obstructive pulmonary disease, 6) insulin-dependent diabetes mellitus, 7) Parkinson’s disease, 8) Alzheimer’s disease or any type of dementia, 9) active cancer treatment (e.g. radiotherapy, chemotherapy), 10) liver and/or kidney disease, 11) epilepsy, 12) current use of five or more prescription medications that affect metabolism or cardiovascular and hemodynamic responses to exercise, 12) use of any ambulatory assistive device.

Research Protocol

Volunteers visited the Physical Activity and Health Laboratory and provided written informed consent. They completed a health history questionnaire, the Physical Activity Readiness Questionnaire (PAR-Q) and the modified NASA-physical activity scale (Scale range: 0-7) (Appendices C-E). Participants then completed a short battery of functional performance (SPPB) tests including: a) balance test - ability to stand with the feet together in the side-by-side, semi-tandem, and tandem positions, b) 8-foot walk, and c) time to rise from a chair and return to the seated position five consecutive times (47). We required participants to achieve a score of 12 (maximal possible score) in order to minimize chances of enrolling participants with any type of lower extremity impairment

(47). Lastly, since this study involved a physical activity protocol, we obtained medical clearance from participant physicians before scheduling their second visit (Appendix H).

During the second visit, participants arrived at the laboratory after fasting for 4 h and refraining from any exercise for at least 12 h. Upon their arrival, they rested in a seated position for a 5-min period, which was followed by resting heart rate and blood pressure measurements. Height and weight measures were also taken and participants were then asked to rest in a supine position for 15 min. A handheld portable indirect calorimetry system, the MedGem Analyzer (Healthe Tech, Inc, Golden, CO), was used to measure resting metabolic rate (RMR). Validity and reliability of the MedGem Analyzer have been reported in adults in a previous study (122).

Next, participants were fitted with three ActiGraph GT3X+ activity monitors, positioned on the dorsal aspect of the dominant wrist, anterior axillary line of the dominant hip, and just above the lateral malleolus of the dominant ankle. The monitors were secured to the body locations using an elastic belt (hip) and two cotton velcro straps (wrist and ankle). The ActiGraph GT3X+ (ActiGraph, LLC, Pensacola, FL) is a lightweight accelerometer-based activity monitor (4.6cm x 3.3cm x 1.5cm, 19g) that measures triaxial acceleration ranging in magnitude from -6 to +6 g. We initialized the monitors to sample triaxial acceleration signals at a sampling rate of 80Hz, which is similar to what is being used in the NHANES activity monitoring study (59,129).

Once fitted with the monitors, participants performed standing still, sitting still, and lying down positions for 30 s each. Participants were then fitted with the Oxycon Mobile indirect calorimetry system (Carefusion, Yorba Linda, CA). This system was programmed to collect expired breath-by-breath data and required participants to wear a facemask and two small units (sensor unit and transmitter unit) mounted on a harness assembly secured to the upper back. The flowmeter and gas analyzers of the Oxycon Mobile were calibrated using a 3-liter air syringe and a known gas mixture (16% O_2 and 4% CO_2). High validity and reliability of this instrument for measuring oxygen consumption in young adults over a range of exercise intensities were reported by Rosdahl et al. (123). After the equipment was properly secured, participants performed one of the two activity routines described in Table 4.1. Each activity was performed for five minutes and participants rested for two to four minutes between activities, allowing for metabolic rate to return to resting levels. Previous studies of this kind have used similar protocols (26,42). At the end of each activity, participants rated their perceived exertion (RPE) using the Borg scale (see Appendix J). The Borg scale has been shown to be valid and reliable in older women aged 75.5 ± 3.8 years (124). Immediately after data collection, accelerometer data were downloaded to a laptop using the software ActiLife 5.0 (ActiGraph Corporation, Pensacola, FL).

Feature Extraction, Data Processing and Algorithm Development

Raw acceleration signals (g) from the three activity monitors (hip, wrist, and ankle) were synchronized and labeled according to the individual activity type (e.g., organizing the room, laundry) and activity category (e.g., household, locomotion). A start and stop record was used to label signals corresponding to the exact times each activity was performed. Accelerometer data not pertaining to any of these activities (e.g. data from rest period) were discarded. Table 4.2. displays the labeling of the individual activities into four different activity categories. Once data were reduced and labeled, a visual inspection was performed to ensure alignment of signals to the corresponding activities.

Examples of acceleration signals for different activities are shown in Figure 4.1. Time-domain features and frequency-domain features (obtained using a Fourier transform) for these acceleration signals were extracted for every 20 s window (Table 4.3).

Acceleration features along with activity labels were the input variables for the following machine learning models: Artificial Neural Network (ANN), Support Vector Machine (SVM), and Random Forest (RF). Previous studies have used these techniques and demonstrated high recognitions rates for activity type (26,32,41,110). A description and illustration of each of these techniques is provided in Table 4.4 and Figure 4.2

Activity classification algorithms for each type of technique were developed using data from individual monitors (e.g., hip alone, wrist alone) and combined monitors (e.g., hip and wrist, wrist and ankle). We developed algorithms using only time-domain features (28 input features) and using a combination of both time- and frequency- domain features (84 input features). For prediction of activity intensity (METs and multiples of RMR), the regression versions of the models were used (e.g., Support vector regression, random forest regression). In order to obtain steady-state activity VO_2, only data from minutes three to five were used for each activity. Similar procedure has been used in previous studies from our laboratory and the literature has also indicated that a two-minute period is usually sufficient for attaining steady-state of VO_2 (26,42,130). To calculate METs for each activity, average steady-state activity VO_2 was divided by 3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ (1). For calculating multiples of RMR ($Mult_{RMR}$), steady-state activity VO_2 was divided by participant resting VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$). Time- and frequency- domain features of the acceleration signals along with MET values or $Mult_{RMR}$ were input features for developing algorithms to predict METs or $Mult_{RMR}$.

Models to estimate locomotion speed were also developed. Speed during the 400 m walk was calculated for each participant using the following equation: Speed ($m\cdot s^{-1}$) = (400/time to complete test in seconds). Time- and frequency- domain features were extracted for acceleration signals from slow treadmill walking (0.8 $m\cdot s^{-1}$) and 400 m walk. These features along with corresponding locomotion speed (0.8 $m\cdot s^{-1}$ and speed during 400 m walk) were used to train the models for estimation of locomotion speed. We chose the former speed because studies have shown that some activity monitoring devices present low accuracy at speeds lower than 0.8-0.9 $m\cdot s^{-1}$ (97,98). Conversely, the speed in the 400 m walk was selected because of its high association with risk for physical disability and mortality (54,61,114).

Statistical Evaluation

Performance of the algorithms for classification of activity type, prediction of activity intensity and locomotion speed were evaluated using a 'leave-one-out' validation approach [6]. The statistics used to test each hypothesis as well as the statistics employed for each exploratory aim are described below.

Hypothesis 1

Our machine learning models would accurately predict activity type in older adults (≥ 80% accuracy).

The accuracy of activity type classification algorithms were tested by calculating percent correct classification for activity group category and individual activity type.

[6] In this technique, all observations except for one is used to train the algorithm and the accuracy of the algorithm is tested on the observation that is left out. This process is repeated as many times as the number of total observations

Hypothesis 2

Our machine learning models would predict activity intensity with similar accuracy observed in previous studies using machine learning models in young adults (bias[7] < ±0.1 MET).

Measurement bias of each algorithm for METs was calculated as the average difference (across all participants) between predicted and actual METs. Similarly, measurement bias of each algorithm for $Mult_{RMR}$ was calculated as the average difference between predicted and actual $Mult_{RMR}$. Linear mixed models[8] were used to determine if predicted METs and $Mult_{RMR}$ were significantly different than actual METs and $Mult_{RMR}$, respectively. Results were considered statistically significant if the 95% confidence interval values did not include zero. Root mean square error (RMSE) values for METs and $Mult_{RMR}$ were calculated by 1) squaring the positive and negative values of the differences between predicted minus actual METs and $Mult_{RMR}$ for each individual, 2) Averaging the squared differences for METs and $Mult_{RMR}$, and 3) calculating the square root of the mean of the squared differences between predicted minus actual METs and $Mult_{RMR}$.

Exploratory Analyses

To determine best monitor placement for activity type recognition, we compared percent correct classification rates of the algorithms trained with data from monitors placed on different sites (hip, wrist, or ankle). To determine best monitor placement for

[7] In the context of this dissertation, measurement bias is defined as the average difference between predicted minus actual values

[8] Linear mixed model is a linear regression analysis that accounts for both random and fixed effects from the predictor variable. It is a particularly useful model for examining linear associations that involve repeated measures on the same variable (131).

activity intensity prediction, we examined bias and RMSE for METs and $Mult_{RMR}$ from the different algorithms.

For examining level of agreement between predicted and actual locomotion speeds for each algorithm, correlation coefficients and coefficients of determination were calculated from Pearson product-moment correlations between actual and predicted speeds. Measurement bias of each algorithm for speed ($m \cdot s^{-1}$) was calculated as the average difference between predicted minus actual speeds. Linear mixed models were used to calculate if the differences were statistically significant. Results were considered statistically significant if the 95% confidence interval values did not include zero.

Ancillary Statistics

A confusion matrix to determine misclassified minutes across the different activity group categories was used to identify the algorithm with the best overall performance. Sensitivity and specificity for classifying the different activity group categories were also calculated for the algorithm with the best overall performance. Sensitivity identifies the number of true events that are correctly classified as such. Conversely, specificity values identify the number of false events that are correctly classified as false events. Measurement bias of the model for time spent in different activity group categories was calculated. Linear mixed models were used to determine if differences were statistically significant. Results were considered statistically significant if the 95% confidence interval values did not include zero.

Software for Developing and Testing the Algorithms

The open source *R statistical software package*, version 3.0.1 - “Good Sport” (www.r-project.org) was used for developing and evaluating the algorithms. Packages ‘nnet’, ‘e1071’, and ‘Random Forest’ were used for applying the ANN, SVM and RF techniques to the development of the algorithms. The following parameters were used for the ANN: 1) 25 hidden units, 2) decay was set to 0.2666667, 3) skip connection layer was allowed, 4) maximum number of weights was set to 10,000, and 4) maximum number of iterations was set to 5000. For the SVM models, a radial-basis kernel was used to minimize the requirement of tuning the function parameters (e.g. penalty terms). In the initial tests, the default parameters performed optimally, thus, we held them constant while developing all SVM models. For the RF models, the number of trees was set to 500.

Results

Participant characteristics are presented in Table 4.5. The final sample was composed of 21 women and 14 men. Participants were healthy, slightly overweight (26.8 ± 4.2 $kg \cdot m^{-2}$) and reported a score of 4.3 ± 1.8 in the NASA Physical Activity Scale (132) (Appendix E). In this scale, the possible scores range from 0 to 7, where each number represents engagement in physical activity of greater physical demand in terms of combined intensity and volume (132). A score of 4 indicates the individual runs less than one mile per week or spends less than 30 min per week in comparable physical activity (132). Thus, participants were relatively inactive based on their NASA Physical Activity Scale score; however, their scores from the physical function tests (SPPB and 400 m walk) indicated absence of lower extremity impairment or mobility disability. All participants were able to hold the three standing positions (side-by-side, semi-tandem, and full-tandem) for ten seconds, and completed the five chair stands in 8.3 ± 1.5 s and the 8-foot walk in 2.5 ± 0.3 s. In addition, all participants completed the 400 m walk and the average speed was 1.17 ± 0.18 $m \cdot s^{-1}$. Reference values for these tests can be found elsewhere (47,54).

Activity type classification

Classification accuracy of the algorithms using only time-domain features were slightly lower (80-91%) than models using both time- and frequency- domain features (87-95%) (Figure 4.3). Additionally, overall percent correct classification of activity type across the entire group revealed that the ANN, RF and SVM models performed similarly for both sets of models (Figure 4.3). The best monitor placement for activity classification was on the wrist followed by ankle, and hip for algorithms using only time-domain features or time- and frequency- domain features (Figure 4.3).

Table 4.6 provides further details on the algorithms performance; it presents percent correct classification by each algorithm for the different activity categories. For the algorithms using hip or ankle data, recognition rates were low for standing, ranging from 0% (RF hip and SVM hip) to 50% (ANN ankle), and modest for recreational activities, ranging from 53% (RF hip) to 69% (ANN ankle). Locomotion, sedentary, and household activity categories were correctly identified 85% to 99% of the time by hip or ankle algorithms. In contrast, the algorithms using wrist data yielded high recognitions rates for standing, ranging from 80% (ANN wrist and SVM wrist) to 82% (RF wrist). For locomotion, sedentary, household, and recreational activities, the recognition rates by the wrist algorithms ranged from 91% to 97%.

The algorithm with the best overall accuracy for classification of activity type was the SVM wrist algorithm. Performance details of the SVM wrist algorithm are provided in Table 4.7 and Figure 4.4. Table 4.7 is a confusion matrix that depicts the performance of the SVM wrist algorithm in classifying activity type[9]. Columns represent predicted activities and rows represent actual activities. The diagonal values (shaded) indicate the number of minutes correctly classified by the model. The misclassification rate for each activity category can be observed within each row. The middle portion of table 4.7 displays the overall accuracy with the 95% confidence intervals (CI). The overall accuracy for each algorithm was obtained as total percent of minutes correctly classified across all participants. The CI represents the lower and upper bound of correct classification for 95% of the participants. The lower panel of table 4.7 presents sensitivity and specificity values of the algorithm for each activity category. According to the confusion matrix, the most significant misclassifications were as follows: the model misclassified 7 min of the 332 min of locomotion as household activity, 11 min of the 481 min of household activity as recreational activity, and 6 min of the 160 min of recreational activity as household activity. Despite these minor misclassification rates, the overall accuracy of the model was 96% (95% CI: 95 to 97%). The model demonstrated high sensitivity (89-99%) and specificity (97-100%) for all activities (Table 4.7). Biases of the SVM wrist algorithm for time spent in the different intensity categories were small and only significantly different than zero for sedentary behavior (0.38 min) (Figure 4.4). Figure 4.5 shows the accuracy of the SVM wrist data model in classifying individual activities. The overall classification accuracy was 78%, varying from 27% for seated posture to 98% for Tai-Chi. The model accuracy was less than 78% for 7 of the 16 activities. Combining data from two or three monitors did not lead to any substantial

[9] Confusion matrices for the other algorithms are presented in Appendix L.

improvement in accuracy level of SVM model for wrist data, and thus, data are not shown.

Activity Intensity Prediction

Table 4.8 displays METs, $Mult_{RMR}$, VO_2, and RPE for each activity. Values for $Mult_{RMR}$ were overall higher than MET values. The highest activity intensity was for simulated bowling, with a MET value of 3.6 ± 0.4 or a $Mult_{RMR}$ value of 4.4 ± 1.2, which is considered moderate intensity PA based on the traditional cutoff point of ≥ 3 METs and < 6 METs (1).

Biases and root mean squared errors (RMSE) for METs predicted by the different algorithms are shown in Table 4.9 (upper panel). The algorithms for prediction of METs were accurate, with biases ranging from 0.00 (RF ankle) to 0.02 METs (ANN hip) and RMSE ranging from 0.51 (RF ankle) to 0.73 METs (ANN wrist). Overall, the RF and SVM algorithms led to lower biases and RMSE than the ANN technique. The only algorithm producing MET estimates that were significantly different than zero – according to the linear mixed model - was the SVM algorithm for processing wrist data. With the exception of this algorithm, all algorithms perform similarly, with no significant influence of monitor placement on MET estimates.

The lower panel of Table 4.9 displays biases and RMSE for $Mult_{RMR}$ predicted by the different algorithms. Biases were significantly different than zero for the RF ankle, SVM hip, SVM wrist, and SVM ankle algorithms. Biases and RMSE values for predicted $Mult_{RMR}$ were higher than for predicted METs.

Estimation of Locomotion Speed

Pearson product-moment correlations indicated that the RF models were superior to the ANN and SVM models in estimating locomotion speed (Table 4.10). The coefficient of determination (R^2) for RF hip, RF wrist, and RF ankle algorithms were 0.71, 0.21, and 0.77 respectively (Figure 4.6). Biases (RMSE) of the RF hip, RF wrist, and RF ankle algorithms for locomotion speed were 0.00 (0.23) mph, -0.03 (0.43) mph, and 0.01 (0.21) mph, respectively. Of the three RF algorithms, only the RF wrist algorithm produced speed estimates significantly different than actual speed.

Discussion

The primary purpose of this study was to develop and evaluate machine learning algorithms for classifying activity type in older adults. We demonstrated that accurate prediction of activity type is possible using machine learning algorithms to process 80Hz data from an ActiGraph GT3X+ activity monitor secured to the wrist, hip or ankle. This result highlights the potential of using machine learning techniques to advance assessment of PA behavior in older adults.

Over the past decade, large-scale studies have relied on cut-point methods to process accelerometer data in older adults (19,20,23). In this age group, the use of machine learning techniques to classify activity type has been restricted to prototypes or monitors used in clinical settings (109). In contrast, studies in younger adults have successfully employed machine learning techniques to process data from commercially available accelerometers, such as the widely used ActiGraph activity monitors (26,27). To our knowledge, our study is the first to develop machine learning algorithms to process ActiGraph data in older adults. In addition, it is one of the few studies to utilize raw acceleration signals from a commercially available accelerometer to develop activity type classification algorithms. Raw acceleration signals provide a large number of data points per second that allow for extraction of both time- and frequency-domain features, increasing the ability to identify signal patterns of activities within short intervals. The successful utilization of 12.8 s intervals by Zhang et al. (41) and 20 s intervals in the present study attest to the fact using short intervals for activity type classification. The algorithms in both studies correctly classified activity type at least 87% of the time. This accuracy level for short windows of time becomes critical in free-living conditions, as most activities are not performed for extended periods. Thus, similar algorithms to those by Zhang et al. (41) and from this study may be helpful in measuring short duration PA in free-living older adults.

A shortcoming of using high sampling rates is the increase in computational burden associated with processing data. To address this issue, the RF and SVM techniques were included in this study as they are efficient in handling large volumes of data (41). The use of the ANN technique with our data was time-consuming and did not lead to greater accuracy compared to RF and SVM models (Figure 4.3). The latter models produced similar results but with less computational burden (RF and SVM: ~25-35 minutes, ANN: >3 hours). Recognition rates of activity type by ANN and RF algorithms ranged from 87% (ANN hip, RF hip) to 94% (ANN wrist, RF wrist), and 87% to 96% by the SVM algorithms (SVM hip, SVM wrist).

Our algorithm performances highlight the possibility of adopting activity type as a PA metric for free-living older adults. Identification of activity type is important because it may help to answer questions on activity level deterioration of those older adults who become physically disabled. For example, how walking time declines in a person who becomes mobility-disabled may be of interest to objectively quantify the magnitude of deterioration. Our algorithms detected locomotion with a recognition accuracy rate greater than 95%. Thus, researchers may use our algorithms to estimate free-living locomotion and examine its association with risk for mobility disability. Moreover, classification algorithms like ours could be used to obtain specific characteristics of locomotion in free-living conditions such as speed. We reported in this study that models using hip or ankle data could provide accurate estimates of locomotion speed (Table 4.10 and Figure 4.6). Both the RF hip and RF ankle algorithms produced locomotion speed estimates that were highly correlated with actual speeds in the 400 m walk (R^2 values of 0.71 and 0.77). These results are of interest because longitudinal studies have reported relationships of locomotion speed with survival time and risks for becoming frail in older adults (54,55).

A topic that has gained attention over the past decade is the effects of sedentary behavior (SB) on health outcomes (38,84–87). Due to this interest, researchers from the PA measurement field have also focused on developing methods to objectively assess SB (39,133). To the best of our knowledge, this study is the first to employ machine learning techniques to quantify SB from an ActiGraph activity monitor in older adults. There is a need to improve assessment of SB in order to better examine its associations with health outcomes. A study by Stamatakis et al. (134) showed that the relationship between SB and cardiometabolic risk differs based on whether a self-report or an objective method is used to quantify SB. This result suggests that the association between SB and health outcomes is partially affected by the accuracy level of the method used to quantify SB. In view of this, machine learning algorithms may also provide more accurate estimates of SB than cut-point methods. Our algorithms, for example, presented correct classification rates of 85% to 97% for sedentary behavior (Table 4.6); more importantly, some of our algorithms (e.g., SVM wrist) are well rounded and can accurately quantify both PA and SB. This is important because Santos et al. (135) reported that objective PA and SB measures are independently associated with functional fitness in older adults. Thus, algorithms such as ours may have fundamental implications to better understand the independent contributions of SB and PA to physical function and risk for physical disability.

Another potential application of machine learning algorithms would be for identification of daily patterns of PA in older adults. Davis and Fox (20) have used accelerometer-based PA estimates to portray daily PA patterns in older compared to younger adults. In their study, average counts per minute for different hours of the day were used to demonstrate diurnal patterns of activity in younger and older adults. In our study, we were able to distinguish between five activity group types in a laboratory setting. Having the ability to do the same in free-living conditions would allow for characterizing the activity types older adults perform during different periods of the day. This would provide additional information to improve our understanding of PA types and patterns in older adults, and could ultimately be used to design physical activity interventions and physical disability preventive strategies.

During the past two decades, the standard placement site for activity monitors has been the hip. This choice predominated because cut-point methods are developed from linear regression equations and, overall, hip acceleration is more linearly related to energy expenditure than acceleration from wrist, ankle or other body location (65,24). With the more recent use of machine learning algorithms, wrist placement has become a good option for use of activity monitors given that accurate PA estimates can be achieved with those algorithms (41). In addition, wrist placement is likely to increase user compliance and also allow for better assessment of upper body motion. The greatest example of the transition to the use of activity monitors on the wrist is the NHANES study, which is currently using the ActiGraph GT3X+ on the wrist at a sampling rate of 80 Hz to collect PA data in a representative sample of Americans (59,129).

Our results support the choice of the wrist as a placement site for accelerometers. Algorithm accuracy was best for the wrist monitor in this study. The SVM algorithm using wrist data achieved the highest recognition rate for activity type with a 96% correct classification rate. The lowest accuracy by the SVM wrist algorithm was for standing, with a correct classification rate of 80% (versus 94-97% for other activity types). A similar level of accuracy for standing was achieved by the RF and ANN wrist algorithms (82% and 80%). In contrast, hip and ankle algorithms resulted in lower classification accuracy for standing, with values ranging from 0-50%. The same trend was observed for recreational activities, where wrist placement produced correct classification rates of 91-94%, whereas hip and ankle placement resulted in recognition rates of only 53-64%. However, it is important to note that machine learning techniques tend to improve when more data are added as inputs. It is possible that additional training data could improve detection of standing and recreational activities. Examining the results from Table 4.6, it appears that algorithms using wrist data were more consistent and accurate than algorithms using hip or ankle data in this study. This may be a result of greater degrees of freedom of arm movement, which in turn results in a broader range of accelerometer signals that are properly captured by our machine learning algorithms.

Even though the main goal of the present study was to predict activity type in older adults, our MET prediction models performed well in estimating energy expenditure (METs). None of the algorithms produced estimates that were significantly different than actual EE estimates. In addition, the biases and root mean square errors (RMSE) of the algorithms were low, indicating the algorithms were accurate for MET prediction (Table 4.9). Staudenmayer et al. (26) also reported low bias and error of their ANN model in predicting METs in younger adults. It is important to note that biases and RMSE from the current models were lower than the values reported by Staudenmayer et al. (26). This is likely because of the differences in activities. In their study, stationary cycling, running, and sport activities (e.g., basketball, racquetball) were the activities that most influenced the magnitude of prediction error. We did not include any of those activities as they are less representative of the PA behavior of the average older adult. Future studies should include these activities, as there is a segment of the older active population who continue to participate in sport activities.

Finally, our results indicate that using resting metabolic rate rather than the standard MET baseline (3.5 ml of $O_2 \cdot kg^{-1} \cdot min^{-1}$) may not be the most appropriate option for training algorithms to predict PA intensity in older adults. Measurement error was larger for $Mult_{RMR}$ than METs in this study. While some studies argue against using METs as a representation of PA intensity, we believe that using individual resting metabolic rate (RMR) creates a greater problem to prediction models (136,137). When measured RMR is used, variability takes place in two ways. First, RMR is different for each individual; and, second, the way metabolic rate varies as a function of multiples of RMR for different activities is also different for each individual. In contrast, when MET is used, the resting metabolic value is assumed to be the same for every individual (3.5 $ml \cdot kg^{-1} \cdot min^{-1}$ of O_2) and variability only occurs in how metabolic rate increases as a function of METs according to different activities for each individual. As a consequence, prediction algorithms developed for use in large-scale studies will likely perform better when using a standard denominator (e.g. 3.5 ml of $O_2 \cdot kg^{-1} \cdot min^{-1}$) for metabolic rate instead of individual RMR.

This study has limitations. We used a protocol comprising structured, fixed time activities, which does not reflect how activities are actually performed in free-living conditions. The participants from this study were relatively healthy and active, which prevent us from generalizing our models to older adults with different characteristics. Finally, testing of model performance was not conducted on a completely independent data sample. This procedure is important to assess if the algorithms perform well when provided with data from different participants and activities. In the study by Freedson et al. (42), the neural network developed by Staudenmayer et al.(26) was tested on a set of data from different participants and some different activities. The accuracy of the neural network in classifying activity type remained high, providing some indication that the model was robust and would be able to classify different activities if it was employed in free-living conditions.

Considering the above limitations, some future research should consider the following factors to enhance the accuracy of the algorithms: 1) train models with additional examples that cover a broader range of intensity, 2) inclusion of participants with different activity or fitness level, 3) testing robustness of algorithms by applying the algorithms to an independent sample, and 4) test models in free-living conditions.

Summary and Conclusions

We hypothesized that our machine learning models would accurately predict activity type in older adults (≥ 80% accuracy, and small bias). In this study, this hypothesis was supported. The machine learning algorithms predicted activity type with correct classification rates of 80% or greater and the biases for time spent in different intensity categories were close to zero. We have also hypothesized that our machine learning models would predict activity intensity with similar accuracy observed in previous studies using machine learning models in young adults (bias < ±0.1 MET). This hypothesis was supported for METs but not for $Mult_{RMR}$. All three SVM algorithms (wrist, hip and ankle) as well as the RF ankle algorithm produced bias significantly different than zero $Mult_{RMR}$.

In this investigation, two exploratory analyses were also conducted. The first exploratory analysis examined the best monitor placement for prediction of activity type and intensity. It was found that wrist placement was superior for prediction of activity type and that prediction of METs was not significantly influenced by monitor placement. In contrast, our algorithms did not predict $Mult_{RMR}$ with the same accuracy and precision as for prediction of METs. All three SVM algorithms and RF ankle algorithm produced estimates of $Mult_{RMR}$ significantly different than actual $Mult_{RMR}$. Additionally, biases and RMSEs were also higher than those for METs. The second exploratory analysis aimed to develop machine learning algorithms to estimate locomotion speed from accelerometer data. It was observed that RF algorithms were superior to ANN and SVM algorithms in estimating locomotion speed. In addition, locomotion speed was accurately estimated by the RF hip and RF ankle algorithms, but not by the RF wrist algorithm. The correlations were high between locomotion speed predicted by the RF hip or RF ankle algorithms and actual speed.

Based on our results, activity type in older adults may be accurately classified from raw acceleration signals collected with an ActiGraph GT3X+ activity monitor. High overall recognition accuracy for activity type can be achieved using ANN, RF, or SVM algorithms for processing hip, wrist, or ankle acceleration data. Our results, however, suggest that higher recognition rate is achieved when using a SVM algorithm to process wrist acceleration data. For prediction of METs, any of the three machine learning techniques produce accurate estimates, which are not significantly influenced by monitor placement. Conversely, the results suggest that using $Mult_{RMR}$ as the measurement unit for PA intensity may result in less accurate and precise predictions by machine learning algorithms. Finally, locomotion speed in older adults can be accurately estimated from acceleration signals using our RF hip and RF ankle algorithms. This may have potential implications for studies examining the associations of locomotion speed with risk for disability. While the results are promising, further testing of our algorithms in free–living conditions is necessary before they are implemented in other studies.

Tables

Table 4.1: Activity routines

Routine 1	Routine 2
Crosswords	Playing cards
Self-care (miscellaneous)	Laundry
Organizing the room	Dusting
Gardening	Vacuuming
Carrying groceries	Slow walk (0.8 $m \cdot s^{-1}$)
400 m walk	400 m walk
Tai-Chi	Playing Bowling

See appendix I for description of activities.
Note: Each participant performed only one of the activity routines. Each activity was performed for 5 min and participants rested for 4 min between activities.

Table 4.2: Categorization of individual activities into activity groups for labeling signals

Activity Group			
Sedentary Behavior	**Locomotion**	**Household/Moving intermittently**	**Recreational**
• Lying down • Sitting • Crossword Puzzles • Playing cards	• Slow walk • 400 m walk • Carrying groceries	• Dusting • Gardening • Vacuuming • Self-care • Laundry • Organizing the room	• Tai-chi • Simulated Bowling

These activity group categories were used to label acceleration signals in order to develop and train the activity type classification algorithms. Acceleration signals were matched to the corresponding activity label based on the start and stop time of each activity.

Table 4.3: Time- and domain- features extracted for training the activity classification algorithms

<table>
<tr><th>Time-domain features</th><th>Frequency-domain features</th></tr>
<tr><td>10th, 25th, 50th, 75th, 90th percentiles of acceleration signals (g)
Mean acceleration (g)
Standard deviation of acceleration (g)</td><td>25th, 50th, 75th, 90th percentiles of signal frequency
Range of frequency distribution
Total signal power
Mean frequency
Dominant frequency
Power of dominant frequency
Second dominant frequency
Power of second dominant frequency
Dominant frequency between 0.6 – 2.5 Hz (df625)
Power of df625
Entropy
Entropy density
Ratio noise/signal</td></tr>
</table>

Each feature is extracted for each of the 3 axes (x,y and z) and also for the composite vector magnitude. g is the abbreviation for g-force (gravitational force). One g corresponds to approximately 9.8 m/s^2, which is the acceleration due to gravity at the earth's surface.

Table 4.4: Description of the machine learning techniques employed to develop activity classification algorithms in the present study

Machine learning technique	Description
Artificial Neural Network (ANN)	Artificial neural networks are computational techniques that mimic biological systems. In an ANN, nodes represent neurons and the links to the different nodes represent neuronal connections (106). The basic components of an ANN are shown in figure 4.2a. The left side of the figure is the input layer, the middle portion shows the hidden layer, and in the far right is the output layer. One or more nodes form each of these layers. The type of neural network used in the present study was a feed-forward multilayer perceptron, which is the most commonly used type of ANN in pattern recognition (106). In a feed forward neural network, information moves unidirectionally from input nodes, through hidden nodes and to the output nodes. An ANN operates as follows: Nodes from the input layer (x_1, x_2, x_n) are linearized (linear transformation) by the hidden layer, which then applies a nonlinear activation function (logistic sigmoid function) to the hidden variables in order to produce the output that can be observed in the output layer. At first, the input variables receive random weights and these weights are then adjusted through *n* cycles of iterations, in which the model minimizes the cost of function $C=[(f(x)-y)^2]$, where *x* is the input feature and *y* is the known variable to be predicted (106). In simple terms, an ANN is a model that comprises multiple layers of linear and nonlinear functions. A good definition of an ANN is that provided by Bishop (106), in which he states: "Thus the neural network model is simply a nonlinear function from a set of input variables $\{x_i\}$ to a set of output variables $\{y_k\}$ controlled by a vector w of adjustable parameters."
Support Vector Machines	Support vector machines are classifiers that find optimal separating decision hyperplanes between classes, implying maximum possible distance between data points belonging to different classes (maximum margin classifier) (Figure 4.2 b). For complex nonlinear functions, SVMs can project data from the original feature space into a hyper-dimensional space (32). With this process, a linear separation can be performed in the hyper-dimensional space. This solution is equivalent to a nonlinear separation in the original feature space (32). In the past 10 years, several studies have used SVMs to classify activity type from accelerometer data (32,41,138). In these studies, SVMs have proved to be a good alternative for accurate and efficient activity recognition.
Random Forest	A random forest consists of a collection of decision tree classifiers (139). A random forest classifier makes a random selection of *n* features from the complete set of features (*N*) for each tree. These *n* features can be termed as the training features. The model then makes the best split on the selected *n* features. The final node of each of the *k* number of trees vote for a given output. The majority vote determines the predicted output (139) (Figure 4.2 c).

See Figure 4.2 for illustration of each technique.

Table 4.5: Participant characteristics

Participant characteristics	
n	35 (21F, 14M)
Age (years)	70.6 ± 5.0
Body Mass (Kg)	76.4 ± 14.4
Height (cm)	168.6 ± 9.9
BMI ($Kg \cdot m^{-2}$)	26.8 ± 4.2
PA score (0 to 7)	4.3 ± 1.8
5 Chair stands – SPPB (s)	8.3 ± 1.5
8 foot walk – SPPB (s)	2.5 ± 0.3
400 m – walk speed ($m \cdot s^{-1}$)	1.17 ± 0.18
Resting VO_2 ($ml \cdot kg^{-1} \cdot min^{-1}$)	3.0 ± 0.6

Values are mean and standard deviation. The score on the balance test from the SPPB is not provided in the table, as all participants were required to hold in each of the standing positions for ten seconds (max score).

Table 4.6: Percent correct classification of the algorithms for each activity group

		Activity Category					
		Loc	**Sed**	**House**	**Rec**	**Stand**	**Overall**
Algorithm	**ANN Hip**	98%	87%	88%	64%	45%	87%
	ANN Wrist	97%	94%	94%	91%	80%	94%
	ANN Ankle	98%	85%	89%	69%	50%	88%
	RF Hip	99%	92%	91%	53%	0%	87%
	RF Wrist	96%	93%	95%	92%	82%	94%
	RF Ankle	99%	89%	92%	61%	40%	89%
	SVM Hip	98%	92%	91%	55%	0%	87%
	SVM Wrist	97%	97%	96%	94%	80%	96%
	SVM Ankle	99%	92%	93%	64%	20%	90%

Values are percent of total time correctly identified for each activity group across all participants. Overall percent correct classification (last column on the right side) is percent correct classification across all activities and participants. Loc: Locomotion, Sed: Sedentary, House: Household, Rec: Recreational, Stand: Standing.

Table 4.7: Confusion Matrix and Sensitivity and specificity for SVM using wrist data

		SVM Wrist Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	322	2	7	1	0
	Sedentary	0	174	4	2	0
	Household	3	5	460	11	1
	Recreational	1	2	6	151	0
	Standing	0	1	0	1	8

Overall accuracy: 96%
(95% CI: 95% - 97%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	95%	96%	91%	89%
Specificity	99%	99%	97%	99%	100%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Table 4.8: VO_2, Multiples of RMR, METs and RPE for each activity

	VO2 ($ml\cdot kg^{-1}\cdot min^{-1}$)	Multiples of RMR (VO_2/RMR)	METs (VO_2/3.5)	RPE
Crossword puzzles	3.5 ± 1.4	1.1 ± 0.1	0.9 ± 0.1	7
Playing cards	4.3 ± 1.9	1.4 ± 0.3	1.2 ± 0.2	7
Laundry	6.9 ± 2.3	2.4 ± 0.5	2 ± 0.4	9
Tai-chi	6.8 ± 2.2	2.4 ± 0.5	2.1 ± 0.3	10
Self-care (miscellaneous)	7.9 ± 3.7	2.6 ± 0.7	2.1 ± 0.6	9
Dusting	7.8 ± 2.9	2.8 ± 0.7	2.4 ± 0.4	9
Gardening	7.8 ± 3.1	2.8 ± 0.7	2.4 ± 0.4	11.5
Vacuuming	9.9 ± 3.1	3.6 ± 0.7	3.1 ± 0.5	11
Organizing the room	10.5 ± 3.5	3.7 ± 0.5	3.2 ± 0.4	11
Slow walk (0.8 $m\cdot s^{-1}$)	10.3 ± 3.3	3.7 ± 1	3.1 ± 0.7	12
400m walk	11.7 ± 3.5	4.1 ± 0.9	3.6 ± 0.6	12
Carrying groceries	11.6 ± 3.6	4.1 ± 0.8	3.5 ± 0.6	11
Simulated Bowling	11.9 ± 4.4	4.4 ± 1.2	3.6 ± 0.4	13

Values for VO_2, Multiples of RMR and METs are mean ± SD. Values for RPE are median values. Note: One MET= 3.5 ml of $O_2\cdot kg^{-1}\cdot min^{-1}$. RMR in this study was 3.01 ± 0.57 ml of $O_2\cdot kg^{-1}\cdot min^{-1}$

Table 4.9: Bias and root mean squared error (RMSE) of each prediction model and monitor placement for prediction of METs and $Mult_{RMR}$

	MET (3.5 ml/kg/min)					
	ANN		RF		SVM	
	Bias	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
Hip	0.02	0.67	0.00	0.52	0.00	0.57
Wrist	0.00	0.73	-0.01	0.54	-0.02	0.55
Ankle	-0.01	0.72	0.00	0.51	-0.01	0.52

	Multiples of RMR#					
	ANN		RF		SVM	
	Bias	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
Hip	-0.02	1.02	0.02	0.86	-0.04*	0.88
Wrist	-0.03	1.14	0.02	0.83	-0.09*	0.89
Ankle	-0.02	1.16	0.05*	0.93	-0.05*	0.96

 # Average resting VO_2 in this study was 3.01 ± 0.57 $ml \cdot kg^{-1} \cdot min^{-1}$
** Significantly different than zero at $p<0.05$*
Bias is the average difference between predicted minus actual values
Values are average across all individuals and activities

Table 4.10: Correlation between predicted and actual locomotion speed, and bias and root mean square error (RMSE) of predicted minus actual locomotion speed

Algorithm	r	R^2	Bias ($m.s^{-1}$)	SD Bias ($m.s^{-1}$)	RMSE ($m.s^{-1}$)
ANN hip	0.64	0.41	-0.02	0.14	0.31
ANN wrist	0.33	0.11	-0.03	0.18	0.51
ANN ankle	0.81	0.66	-0.01	0.10	0.25
RF hip	0.84	0.71	-0.01	0.10	0.23
RF wrist	0.46	0.21	-0.04	0.16	0.43
RF ankle	0.88	0.77	-0.01	0.09	0.21
SVM hip	0.72	0.52	-0.02	0.12	0.29
SVM wrist	0.29	0.09	-0.04	0.18	0.45
SVM ankle	0.86	0.74	-0.02	0.10	0.24

Correlation values (r and R^2) were obtained using Pearson product-moment correlations between individuals' average predicted speed and individuals' average actual speed. Note that average values for each individual were used in order to preserve independence between predicted variable (predicted speed) since prediction was done for every 20 s. Using every predicted value would lead to overestimation of the strength of correlation and would also violate assumptions for using Pearson correlation. Bias is the average difference between predicted minus actual locomotion speed calculated from all sample. RMSE is the square root of the average of the squared differences between predicted minus actual locomotion speed.

Figures

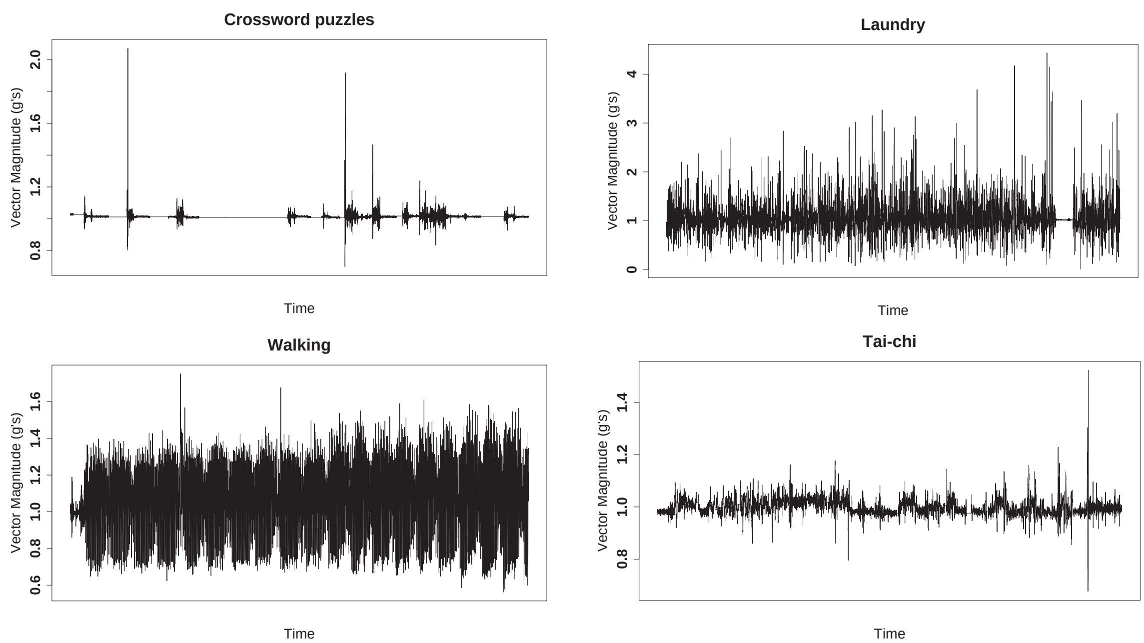


Figure 4.1: Vector magnitude from wrist acceleration signals (g) for 4 different activities.

Vector magnitude (VM) is calculated as VM=$\sqrt{x^2+y^2+z^2}$, where x is vertical acceleration, y is anteroposterior acceleration, and z is mediolateral acceleration. Time between two consecutive data points in each of the plots is equivalent to 1/80 s since data were collected at 80 Hz. For the 5-min duration of each activity, 24000 data points were collected.

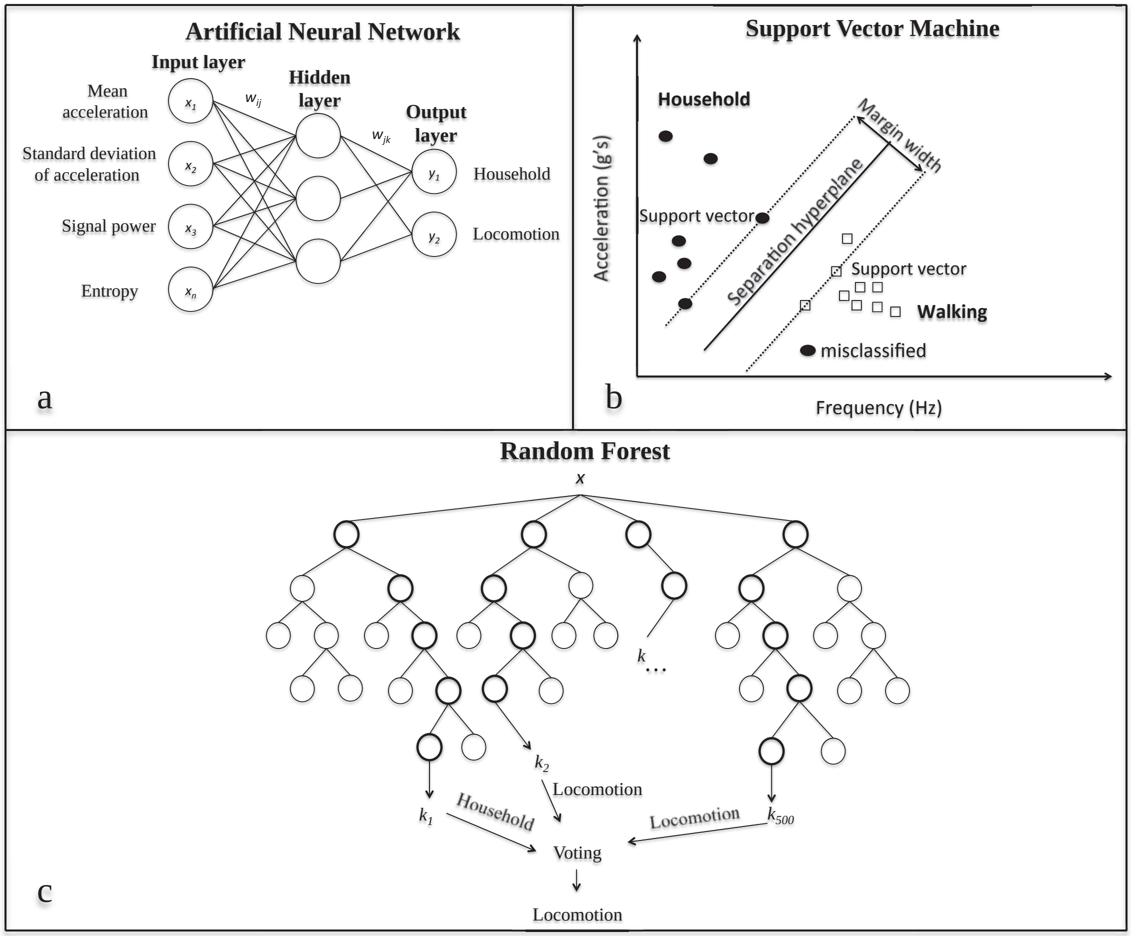


Figure 4.2: Illustration of the machine learning techniques used to develop and train the activity classification algorithms in the current study.

A description of each of these machine learning techniques is provided in Table 4.4. Note that each of the techniques attempts to identify patterns but in different ways.

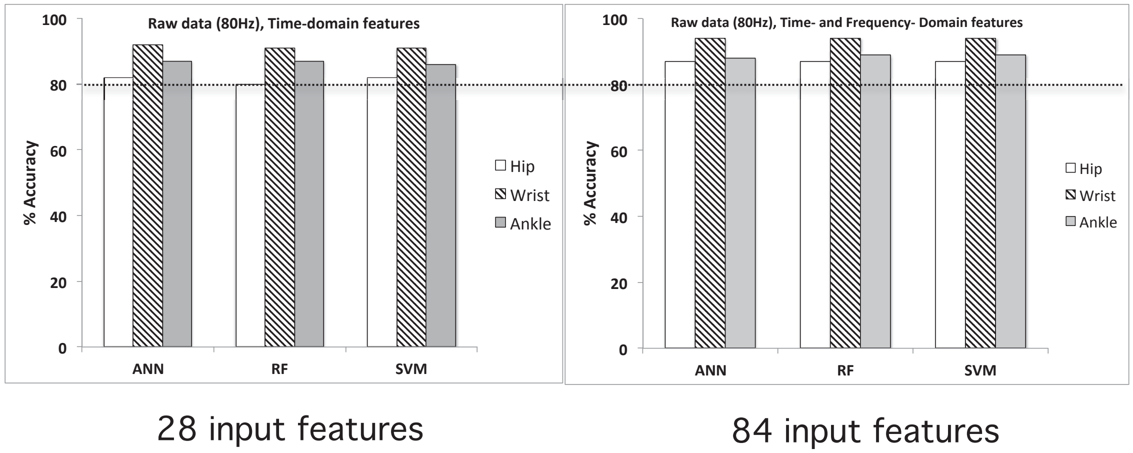


Figure 4.3: Performance of ANN, RF and SVM models based on hip, wrist and ankle accelerometer signals.

Left panel shows performance of algorithms using only time-domain features. Right panel shows performance of algorithms using both time- and frequency- domain features. Note that the two sets of algorithms utilize different number of training input features. The *y-axis* of each figure is overall percent correct classification of activities for combined data from all participants. The *x-axis* displays the bars for ANN, RF, and SVM algorithms. Each bar denotes a different monitor placement (see legend). The dotted line indicates 80% percent correct classification. This was the accuracy level (minimum) we aimed for in this study.

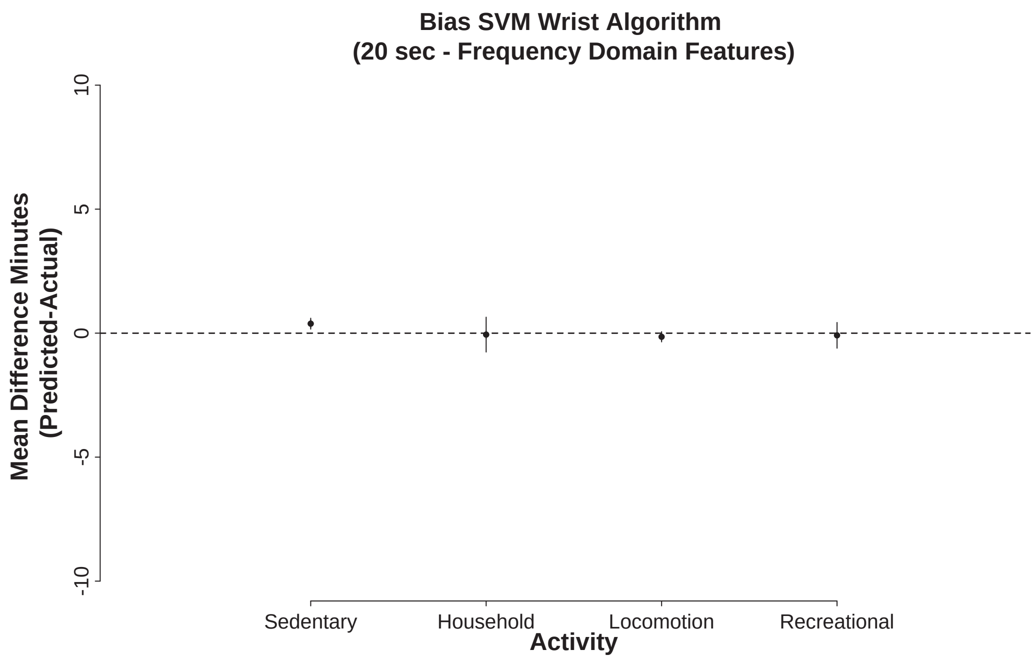


Figure 4.4: Bias (minutes) of SVM Wrist Algorithm for each activity group category

The *y-axis* displays mean difference in minutes (bias) between predicted minus actual time spent in different activity categories. The *x-axis* displays the different activity categories used in the current study. Note that standing is not included in this analysis as only ten minutes were available for this activity. Black dots are mean values and error bars are 95% confidence intervals (CI). Linear mixed models indicated that only estimates for sedentary activity were significantly different than zero. Observe that lower bound of 95% CI does not cross zero for sedentary activity. All other estimates were not significantly different than zero. Values are relative to 35 min of activity (Sedentary: 5 min, Household: 15 min, Locomotion: 10 min, Recreational: 5 min). These durations were the same for all participants.

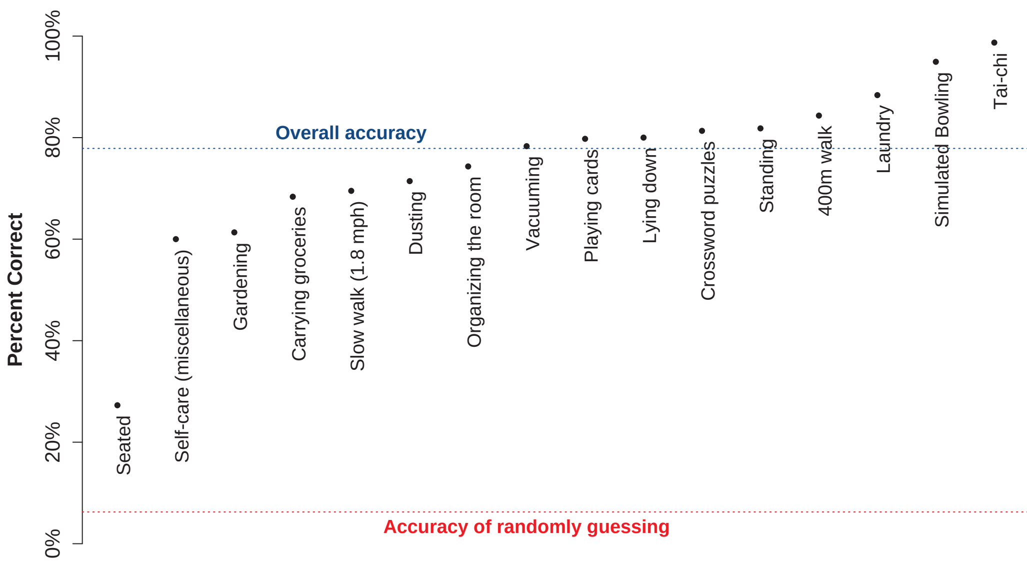


Figure 4.5: Performance of SVM wrist algorithm for prediction of individual activities

The *y-axis* displays percent of total time for each activity (group values) that is correctly classified by the algorithm. The *x-axis* shows the individual activities. Overall accuracy line depicts the recognition rate of the algorithm across all activities. Accuracy of randomly guessing is the probability of the model in correctly classifying an event if no method was employed (literally ‘guessing’). In this study, this chance would be 1/16, where the denominator is the total number of activities.

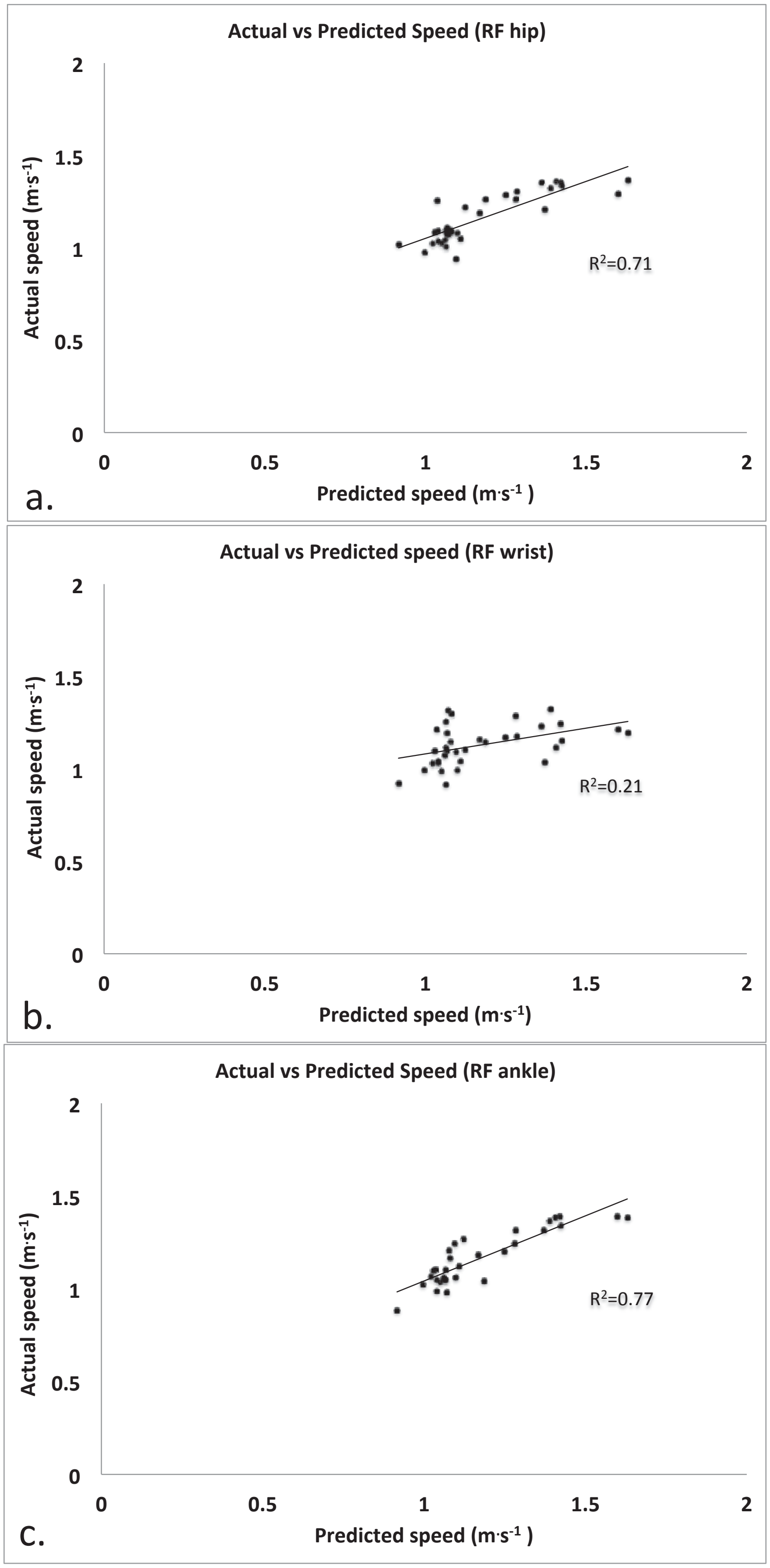


Figure 4.6: Prediction of locomotion speed by RF algorithms using (a) hip, (b) wrist, and (c) ankle data.

The *y-axis* is the average actual speed calculated from the 400 m walk. The *x-axis* is the average speed predicted by the algorithm. Each data point is the average predicted speed for each participant.

CHAPTER V

VALIDATION OF ACCELEROMETER-BASED ACTIVITY CLASSIFICATION ALGORITHMS IN FREE-LIVING OLDER ADULTS

Abstract

Purposes: 1) To compare activity type recognition rates of random forest and support vector machine algorithms trained on laboratory (RF_{Lab} and SVM_{Lab}) versus free-living accelerometer data (SVM_{FL} and RF_{FL}) in free-living older adults, 2) to examine the correlation between locomotion speed predicted by an algorithm developed in Study 1 (RF_{speed}) and speed in the 400 m walk test. **Methods:** Fifteen participants from Study 1 wore three ActiGraph GT3X+ activity monitors (dominant hip, wrist, and ankle) in free-living conditions and were directly observed for 2-3 h. The RF_{Lab} and SVM_{Lab} algorithms were applied to hip, wrist, and ankle accelerometer data for classification of activity type in 20 s intervals. Free-living accelerometer data were used to train SVM_{FL} and RF_{FL} algorithms for classification of activity type. Direct observation data were used to compute percent correct classification for activity type by the different algorithms. The best algorithm was used to predict locomotion time and the RF_{speed} algorithm was applied to predict locomotion speed. A Pearson product-moment correlation was used to determine the association of predicted with actual locomotion speed (400 m walk speed). **Results:** Overall classification accuracy using 20 s intervals for the lab-based algorithms was between 49% (wrist) to 55% (ankle) for the SVM_{Lab} algorithms, and 49% (wrist) to 54% (ankle) for RF_{Lab} algorithms. Overall classification accuracy of SVM_{FL} and RF_{FL} algorithms ranged from 58% (wrist) to 69% (ankle) and from 61% (wrist) to 66% (hip and ankle), respectively. Using 30 s intervals improved classification accuracy up to 71%

(SVM_{FL} ankle). Significant improvements in classification accuracy were observed for RF_{FL} hip, RF_{FL} wrist, and RF_{FL} ankle algorithms (76%, 70%, and 76%) when three activity type categories were used: sedentary behavior, moving intermittently and locomotion. The highest predicted locomotion speed attained by the free-living older adults was moderately correlated (r=0.55) to speed in the 400 m walk. **Conclusions:** The activity classification algorithms trained with free-living data were more accurate in predicting activity type in this study compared to laboratory algorithms. Machine learning algorithms may be useful in predicting locomotion speed in free-living older adults.

Introduction

In study 1, we developed classification algorithms to detect activity type from wearable accelerometer data in older adults. The algorithms were accurate in identifying activity types in laboratory conditions (87% - 95% accuracy). However, the performance of these algorithms in free-living settings is unknown. This is important because the final application of these algorithms is the measurement of physical activity (PA) in the free-living environment. In the laboratory, the methods to develop and validate lab-based algorithms are highly controlled to first establish proof of concept. In natural settings, activities are not performed in known and fixed time intervals and different people perform these less constrained activities differently. Therefore, accuracy in laboratory settings does not ensure that activity type classification algorithms will perform well under free-living conditions.

Two studies demonstrated that recognition rate for activity type drops significantly when algorithms trained with laboratory accelerometer data are tested on free-living accelerometer data. In a study by Foerster et al. (62), a machine learning algorithm yielded a recognition rate of 100% in laboratory conditions but was only able to recognize 67% of the activities in free-living conditions. Similarly, Ermes et al. (63) reported that an algorithm trained on both laboratory and free-living accelerometer data recognized activity type with 89% accuracy in free-living conditions; however, when the algorithm was only trained on laboratory accelerometer data, recognition rate dropped to 72%. To date, studies that were specifically conducted in older adults have only developed and tested validity of machine learning algorithms in clinical settings or during pre-determined activity routines (109). These studies used prototypes of accelerometers or activity monitors that are typically used in clinical settings (109). Our study (Study 1) was the first to use accelerometer data from a widely used activity monitor (ActiGraph™ GT3X+) to develop activity classification algorithms for older adults. It is essential to test the validity of these algorithms in free-living conditions and to adjust the prediction models to such conditions. Refinements of machine learning algorithms have been conducted in previous studies and were important to achieve acceptable accuracy for use of the algorithms in free-living settings (63,133).

For example, recent work in our laboratory confirmed the difficulty in directly applying a machine learning method developed in the laboratory for prediction of energy expenditure in natural settings (133). The neural network developed by Staudenmayer et al. (26) was compared to direct observation in free-living young adults and it performed poorly in predicting MET-hours and time spent in different activity intensity categories. To refine the algorithm, the ‘Sojourn method’, which is a decision tree model, was first applied to the monitor data to identify sedentary behavior and activity bouts. The neural network algorithm was then applied to the bouts of data labeled by type of activity. Compared to the original neural network, the two-step ‘sojourn’ method substantially improved estimations of MET-hours and time spent in different intensity categories (133). For classification of activity type, the study by Ermes et al. (63) is the best example of refinement of an algorithm to classify activity type from accelerometer data in free-living conditions. In their study, recognition rate of a hybrid machine learning algorithm (decision tree and artificial neural network) increased by ~17% when the algorithm was trained on both laboratory and free-living accelerometer data as opposed to only training on laboratory data. Both studies suggest that algorithms developed in the laboratory require refinement for free-living physical activity (PA) applications.

We are in the early stages of detecting activity type with wearable devices in free-living conditions. It is essential to determine how these algorithms perform in natural settings so that they may be used in future PA intervention and surveillance research to understand the relationship between activity type and physical function in older adults. For example, locomotion time and speed could be quantified in the free-living environment, and these locomotion variables can be correlated with selected health or functional outcomes. In study 1, we developed algorithms to predict locomotion speed from accelerometer data in older adults. Accuracy of the Random Forest algorithms for predicting locomotion speed from hip and ankle data was high in laboratory conditions. Nevertheless, similar to the algorithms that predict activity type, these locomotion speed prediction algorithms need to be first tested in free-living settings, as their accuracy in such conditions is unknown.

Thus, the purposes of this study were to: 1) test the accuracy of our lab-based algorithms in detecting activity type in free-living older adults, 2) develop and evaluate algorithms using free-living accelerometer data, and 3) to correlate estimates of free-living locomotion speed predicted by one of our lab-based algorithms with performance in the 400 m walk, which is a well-established field test to assess physical function in older adults.

Hypotheses

Hypothesis 1: The machine learning algorithms developed in Study 1 will classify activity type from accelerometer data in free-living older adults with similar accuracy as previous studies (~70% accuracy) (62,63).

Hypothesis 2: Algorithms developed with free-living accelerometer data will classify activity type in free-living older adults more accurately than lab-based algorithms developed in Study 1.

Exploratory Analysis

Locomotion speed predicted by a machine learning algorithm will be correlated to speed in the 400 m walk from Study 1. The purpose of this analysis is to examine if locomotion speed predicted by machine learning algorithms may be used as a marker of physical function in free-living conditions. This will be the first study examining this association. If significant results are found, it may indicate that measuring free-living locomotion speed using machine learning algorithms is an alternative to assessing speed during a 400 m walk.

Methods

Participants

The selection of participants for the present investigation was based on a convenience sample. Fifteen older adults (6 men and 9 women) who participated in Study 1 were invited to take part in this study. These participants were previously screened for Study 1 and did not present with any of the following conditions: 1) age <65 or >85 years, 2) diagnosed heart disease, 3) myocardial infarction or stroke in the past year, 4) congestive heart failure, 5) chronic obstructive pulmonary disease, 6) insulin-dependent diabetes mellitus, 7) Parkinson's disease, 8) Alzheimer's disease or any type of dementia, 9) active cancer treatment (e.g. radiotherapy, chemotherapy), 10) liver and/or kidney disease, 11) epilepsy, 12) current use of five or more prescription medications that affect metabolism or cardiovascular and hemodynamic responses to exercise, 12) use of any ambulatory assistive device.

Participants read and signed the written informed consent form, which was approved by the Institutional Review Board from the University of Massachusetts Amherst. All the participants had previously (Study 1) completed the Short Physical Performance Battery test (SPPB) and achieved a score of 12, indicating they did not have lower extremity impairments. The SPPB is composed of three components: a) balance test - ability to stand with the feet together in the side-by-side, semi-tandem, and tandem positions, b) 8-foot walk, and c) time to rise from a chair and return to the seated position five consecutive times (47). Another physical function test completed by the participants was the 400 m walk, where they walked 10 laps at their regular speed over a 40 m course. Participants also completed the following questionnaires: 1) Health history questionnaire, 2) Physical activity readiness questionnaire (PAR-Q), and the 3) NASA physical activity scale (Scale range: 0 to 7) (Appendices C-E). Participant characteristics are presented in table 5.1. After consenting to take part in Study 2, participants were scheduled to be directly observed in their free-living environment for a 2-3 hour time block.

Instrumentation

Direct Observation System

A Personal Digital Assistant (PDA) programmed for *continuous focal sampling direct observation* (CFS-DO) (*The Observer*®; Noldus Information Technology, Wageningen, The Netherlands) was used to code the activities performed by the participants in the free-living environment. The PDA and CFS-DO software were used to capture two PA variables:

1) Activity type - A menu of activities for the PDA was created by two of the authors (PSF and JES). The menu included activities that are commonly performed by older adults (Table 5.2). The selection of appropriate activities was also based on time use data for older adults (12).
2) Activity duration: A 1 s record interval was used to record the activities.

Activity Monitors

Three ActiGraph GT3X+ (ActiGraph Inc, Pensacola, FL) activity monitors were synchronized to the same computer and initialized using the ActiLife 5 software to collect raw triaxial acceleration signals (g) at a sampling rate of 80 Hz. Prior to the scheduled DO session, the observers assembled monitors onto an elastic belt (hip unit) and two elastic straps (wrist and ankle) in preparation for meeting with the participant.

Observers

Three observers were trained to use the PDA and the continuous focal sampling software. They received instructions on coding activity type and duration during face-to-face training sessions and group discussion meetings. At the end of the training period, the observers completed a test to examine inter-observer reliability. The test consisted of coding activity type and duration of twenty activity video clips. The *Cohen's kappa coefficient* for inter-observer agreement for coding activity type and duration was 0.89.

Direct Observation

Observation sessions were scheduled for daytime (9 am to 5 pm). A single block of 2-3 hours of DO was carried out for each participant. The observers met the participants at the pre-determined time and location. Before starting the DO session, observers assisted participants with placement of the monitors (dominant wrist, hip and ankle). Participants were instructed to perform their daily routine as normally as possible. Observers started the DO session and recorded type and duration of the activities performed by the participants. There were no instructions as to how they should perform activities and we did not request participants to engage in any particular activity after the observation session was started.

Accelerometer Data Processing and Statistical Evaluation

Accelerometer data were processed using activity type classification algorithms developed in Study 1. We applied the Support Vector Machine and Random Forest algorithms[10] (SVM_{Lab} and RF_{Lab}) to wrist, hip, and ankle accelerometer data to classify five activity group categories: 1) Standing, 2) Sedentary, 3) Household, 4) Locomotion, and 5) Recreational activity. We also developed new algorithms using free-living accelerometer data with direct observation labels (Figure 5.1 and Table 5.2). Similar to Study 1, we extracted time- and frequency- domain features from the acceleration signals (Figure 4.3 from Study 1), which were used with the direct observation labels to train SVM and RF algorithms to classify activity type from free-living accelerometer data (SVM_{FL} and RF_{FL}). All analyses were conducted using the open source *R statistical software package*, version 3.0.1 - “Good Sport” (www.r-project.org). The r-packages

[10] The algorithms were developed for classification of 20 s intervals (see study 1).

‘e1071’ and ‘Random Forest’ were used for applying and developing SVM and RF algorithms for classification of activity type. Statistics for testing each of the hypothesis and exploratory aims are presented below.

Hypothesis 1

The machine learning algorithms developed in laboratory will classify activity type from accelerometer data in free-living older adults with similar accuracy as previous studies (~ 70% accuracy) (62,63)

Direct observation was used as the criterion method to calculate percent correct classification of activity type by the SVM_{Lab} and RF_{Lab} algorithms. Overall percent correct classification by the algorithms was calculated with the following equation: *% correct = (minutes in different activity categories that are correctly identified by the algorithm ÷ total minutes in different activity categories assessed by DO) x 100.* To calculate percent correct classification by the algorithm for a specific activity category (e.g. locomotion), we modified the equation as follows: *% correct for ‘locomotion’ = (minutes in locomotion that are correctly identified by the algorithm ÷ total minutes in locomotion assessed by DO) x 100.*

Hypothesis 2

Algorithms developed with free-living accelerometer data will classify activity type in free-living older adults more accurately than lab-based algorithms developed in Study 1

A leave-one-out cross validation technique[11] was employed where direct observation was used as the criterion method for calculating percent correct classification of activity type by the SVM_{FL} and RF_{FL} algorithms. Classification accuracies of the new algorithms were compared to those of the lab-based algorithms. The same window length (20 s) and five activity groups from Study 1 were used to categorize free-living accelerometer data.

Ancillary Statistics

Ancillary statistics were used to further evaluate the performance of the newly developed SVM_{FL} and RF_{FL} algorithms. In addition to the models for prediction of five activity categories, models were trained to predict three activity groups: 1) Sedentary/Standing, 2) Moving intermittently (Household and Recreational activity combined), and 3) Locomotion. Classification accuracy was tested using classification intervals varying from 5 to 30 s. Sensitivity and specificity analyses were conducted for the SVM_{FL} or RF_{FL} algorithms with the highest accuracy for classification of both five and three activity categories. Sensitivity identifies the number of true events that are correctly classified as such. Specificity values identify the number of false events that are correctly classified as false events. Bias and 95% confidence intervals (min) for the classification of time spent in both five and three activity groups for the algorithms with the best performance were determined.

[11] In this technique, all observations except for one is used to train the algorithm and the accuracy of the algorithm is tested on the observation that is left out. This process is repeated as many times as the number of total observations.

Locomotion Prediction

Locomotion is one of the most important activities performed by older adults. Therefore, the free-living algorithm (RF_{FL} or SVM_{FL}) that produced the highest recognition rate for locomotion was used to predict locomotion time for the participants. A paired *t-test* was used to identify differences between estimated (by the RF_{FL} or SVM_{FL}) and actual locomotion time (obtained from DO). Level of significance was set at $p<0.05$. For those instances where predicted locomotion time was not significantly different than actual locomotion time, speed was predicted using an RF algorithm (RF_{speed}) developed in Study 1. Criterion measure for locomotion speed was obtained from the 400 m walk test performed by the participants in Study 1. Participants completed 10 laps of a 40 m course at their habitual walking speed. Average speed ($m \cdot s^{-1}$) during the 400 m walk was calculated by the following equation: time in seconds to complete the test ÷ 400 m. The 400 m walk test is widely used in clinical settings for physical function assessment in older adults (2).

Exploratory Analysis

To examine if locomotion speed predicted from accelerometer data in free-living conditions is related to speed in the 400 m walk, Pearson product-moment correlations were calculated between: 1) average free-living locomotion speed[12] and 400 m walk average speed, and 2) maximum free-living locomotion speed and 400 m walk speed.

---

[12] Each individual could have more than one value for predicted locomotion speed, as some of them may have performed more than one bout of locomotion. In order to avoid violation of the assumption of independence between measures required to use the Pearson product-moment correlation, only the average value for predicted locomotion speed was used for each individual. Using every single data point would lead to overestimation of the relationship between predicted and actual locomotion speed.

Results

Participants

Participant characteristics are presented in Table 5.1. Overall, participants were slightly overweight (BMI = 26.0 ± 4.3 $kg \cdot m^{-2}$) and reported a score of 4.6 ± 1.7 in the NASA physical activity scale (132) (Appendix E). In this scale, the possible scores range from 0 to 7 and indicate crescent engagement in physical activity of greater intensity and volume combined (132). A score of 4 indicates the individual runs less than one mile per week or spends less than 30 min per week in comparable physical activity (132). Therefore, participants in this study were relatively inactive; however, the scores on the physical function tests (SPPB and 400 m walk) indicated absence of lower extremity impairment or mobility disability. Participants completed the five chair stands in 8.7 ± 1.2 s and the 8-foot walk in 2.5 ± 0.4 s. The average locomotion speed during the 400 m walk was 1.2 ± 0.2 $m \cdot s^{-1}$. Values for the SPPB balance tests are not presented because all participants were able to hold the three standing positions (side-by-side, semi-tandem, and full-tandem) for ten seconds, which is the maximal score for the test (47).

Direct Observation

Participants were observed for an average of 118.1 ± 19.0 min. Participants spent 9.4 ± 19.5 min in recreational activities, 22.6 ± 12.2 min in household activities, 24.3 ± 30.7 min in locomotion, 24.4 ± 13.8 min in standing position, and 33.5 ± 18.7 min in sedentary behavior. Private time, which was requested when participants did not want to be observed, only accounted for 3.8 ± 6.8 min of total observation time. Results on classification accuracy of the algorithms are based on these durations.

Classification of 5 Activity Groups

Classification accuracy of lab-based SVM algorithms in free-living conditions was 49%, 49%, and 55% for SVM_{Lab} wrist, SVM_{Lab} hip and SVM_{Lab} ankle algorithms (Figure 5.2a). Classification accuracy of lab-based RF algorithms was 49%, 51% and 54% for RF_{Lab} wrist, RF_{Lab} hip, and RF_{Lab} ankle algorithms (Figure 5.2c), respectively. The new SVM algorithms developed with free-living data (SVM_{FL}) performed substantially better than the SVM_{Lab} algorithms with accuracy rates of 58%, 64%, and 69% for the SVM_{FL} wrist, SVM_{FL} hip and SVM_{FL} ankle algorithms, respectively (Figure 5.2b). Similarly, accuracy of the new RF algorithms (RF_{FL}) increased to 66%, 61%, and 67% for RF_{FL} wrist, RF_{FL} hip and RF_{FL} ankle algorithms (Figure 5.2d). Table 5.3 displays performance of lab-based and free-living RF and SVM algorithms for each activity group. The RF_{Lab} and SVM_{Lab} algorithms performed extremely poorly for standing (accuracy range: 0-1%) and recreational activity (13-26%), poorly for locomotion (33-52%), and reasonably well for sedentary behavior (62-79%) and household activity (71-87%). Overall, the algorithms trained with free-living data (SVM_{FL} and RF_{FL}) improved the detection for standing (10-52%), recreational activity (20-41%), locomotion (65-80%), and for sedentary behavior (75-87%). However, there was a small decrease in accuracy for household activity (63-73%).

The correct classification rate of the SVM_{FL} wrist, SVM_{FL} hip, and SVM_{FL} ankle algorithms increased to 59%, 65%, and 71% when window length was increased to 30 s (Table 5.4a). Reducing classification interval to 10 s or 5 s resulted in consistent reduction in correct classification rate of the SVM_{FL} algorithms (Table 5.4a). For 30 s classification interval, the SVM_{FL} algorithms were poor in classifying standing and recreational activity. The SVM_{FL} hip, SVM_{FL} wrist, and SVM_{FL} ankle correctly classified standing only 45%, 10% and 53% of the time. For recreational activity, classification accuracy was 22%, 21% and 40% for SVM_{FL} hip, SVM_{FL} wrist, and SVM_{FL} ankle. Accuracy for the other activity groups ranged from 66% to 87% (Table 5.4a).

For the RF_{FL} algorithms, increasing classification intervals to 30 s resulted in improved classification accuracy for RF_{FL} hip and RF_{FL} ankle algorithms (7% and 3%) and reduction in classification accuracy for RF_{FL} wrist algorithm (-3%) compared to 20 s classification interval (Table 5.4b). Reducing classification interval to 10 or 5 s resulted in lower accuracy for all three RF_{FL} algorithms (Table 5.4b). The RF_{FL} algorithms accuracy was poor for recreational activity and standing. For 30 s classification interval, recreational activities were correctly classified only 25%, 24%, and 39% of the time by the RF_{FL} hip, RF_{FL} wrist, and RF_{FL} ankle algorithms (Table 5.4b). Similarly, standing was correctly classified only 40%, 10%, and 50% of the time by the RF_{FL} hip, RF_{FL} wrist, and RF_{FL} ankle algorithms (Table 5.4b). Table 5.5 shows the confusion matrix in addition to sensitivity and specificity values for the RF_{FL} ankle algorithm (30 s classification interval), which exhibited the highest correct classification rate (70%). The lowest accuracy of the RF_{FL} ankle algorithm was for recreational activity. The algorithm correctly classified only 52 min of recreational activity and confused 22, 20, 24 and 16 min as household activity, locomotion, sedentary behavior and standing, respectively. The highest algorithm accuracy was for sedentary behavior with 406 min correctly classified and only 66 min misclassified as other activity groups (Table 5.5). Sensitivity of the RF_{FL} ankle algorithm for the different activity groups varied from 28% (Standing) to 86% (Locomotion). Specificity of the algorithm ranged from 85% (Standing) to 94%

(Locomotion) (Table 5.5). Figure 5.3 shows the bias of the RF_{FL} ankle algorithm for time spent in each of the five activity categories. The biases for the five activity categories were not statistically significant, ranging from -0.6 ± 1.4 min (95% CI: -3.3 to 2.1 minutes) for standing to -2.7 ± 2.1 min (95% CI: -6.9 to 1.4).

Classification of 3 Activity Groups

Since the RF_{FL} algorithms were more consistent than the SVM_{FL} algorithms for classification of five activity groups (Tables 5.3 and 5.4), the RF_{FL} algorithms were retrained for classification of three activity groups in order to test if accuracy could be further improved. The activity groups were 1) Sedentary/Standing, 2) Moving inter (combination of Household with Recreational activity), and 3) Locomotion.

Accuracy of the newly developed RF_{FL} algorithms improved when classifying only 3 activity groups (Table 5.6). The algorithms performed the best with a 30 s classification interval. The RF_{FL} hip and RF_{FL} ankle algorithms exhibited an overall correct classification rate of 76% while the RF_{FL} wrist algorithm produced an overall correct classification rate of 70%. Reducing classification interval led to a reduction in correct classification rate for all algorithms (Table 5.6). Since both RF_{FL} hip and RF_{FL} ankle algorithms using a 30 s classification interval presented similar overall correct classification rate, the confusion matrix for the RF_{FL} hip algorithm is shown because of the consistency across the three different activity groups (Table 5.7). The algorithm misclassified 126 min (24%) of a total of 535 min of ‘Moving intermittently’ as ‘Sedentary/Standing’, 78 min (23%) of a total of 338 min of ‘Locomotion’ as ‘Moving intermittently’, and 124 min (17%) of a total of 735 min of ‘Sedentary/Standing’ as ‘Moving intermittently’ (Table 5.7). Sensitivity values of the algorithm for ‘Moving intermittently', 'Locomotion', and 'Sedentary/Standing' were 65%, 85% and 82%, respectively and specificity values of the algorithm for these activity categories were 83%, 94% and 86% (Table 5.7). Estimated bias for locomotion was -2.3 min (95% CI: -6.6 to 2.0 min), while bias for moving intermittently and sedentary/standing were 1.9 min (95% CI: -5.6 to 9.3 min) and 0.5 min (95% CI: -4.0 to 5.0 min) (Figure 5.4). None of the estimates were significantly different than actual values.

Free-living Locomotion

Figure 5.5 shows time spent in locomotion according to the RF_{FL} hip algorithm and direct observation (actual) for each individual. The RF_{FL} hip algorithm was accurate in predicting time spent in locomotion for 13 of the 15 participants (bias: 1.3 ± 1.75 min). The algorithm was considerably inaccurate for participants 11 and 12, overestimating locomotion time by 16 and 10 minutes compared to differences of up to 6 min for the other participants. The paired t-test indicated significant differences between predicted and actual time spent in locomotion when all participants were included in the analysis. When participants 11 and 12 were removed from the analysis, the differences were no longer significant.

Free-living Locomotion Speed versus 400 m Walk Speed

Based on Figure 5.5, we eliminated data from participants 11 and 12, for whom the estimates were inaccurate, and from three other participants who had 0 min of walking. The RF_{speed} hip prediction algorithm was then applied to the remaining 10 participants. The correlation between average free-living locomotion speed and 400 m walk speed was weak (r=0.22) (Figure 5.6a) whereas the correlation between maximum free-living locomotion speed and 400 m walk speed was moderate (r=0.55) (Figure 5.6b).

We further examined participants with at least 10 min of locomotion. This cutoff point was selected based on the PA guidelines for Americans, in which the minimum duration of a meaningful PA bout is 10 min (140). Only participants 2, 3, 5, 10, 12, and 13 performed at least 10 min of locomotion (Figure 5.5). However, classification of locomotion was poor for participant 12. Therefore, we plotted individual values for average free-living locomotion speed and 400 m walk speed (Figure 5.7 top), and maximum free-living locomotion speed and 400 m walk speed (Figure 5.7 bottom) for the remaining 5 participants. Except for participant 2, in both figures the algorithm indicated that free-living locomotion speed follows the same trend as speed in the 400 m walk test, meaning that the predictions were able to rank individuals from lowest to highest free-living locomotion speed.

Discussion

In this study, we tested the accuracy of our lab-based activity type classification algorithms (Study 1) in free-living older adults. New algorithms were developed using free-living accelerometer data and their accuracy and precision were tested. Our lab-based algorithms performed poorly in free-living conditions. Conversely, accuracy of the free-living algorithms was higher but still not sufficiently accurate for assessment of activity type in natural settings.

The poor performance of our lab-based algorithms was somewhat expected as our group previously reported similar trends for prediction of free-living activity ‘intensity’ (133). This demonstrates that high accuracy in laboratory settings does not translate into high accuracy in free-living conditions. High accuracy of activity type classification algorithms in laboratory studies is observed because these studies have used protocols involving pre-defined activity routines and fixed duration of activities (26,27,41,128). These characteristics make classification easier than in real world situations. In addition, lab-based studies usually test classification accuracy of algorithms using classification intervals (e.g., 1 min) that are defined as a function of the pre-set activity duration (e.g., 5 min, 6 min). Therefore, an algorithm that classifies activity type for 1-min intervals will have a perfect match of five classification events for a 5-min activity in the lab. Activities in free-living conditions are not performed in known intervals and this may lead to substantial degradation of algorithm accuracy in those conditions.

Furthermore, validity of the algorithms is usually tested on the same activities from which they are developed (26,27,41). A leave-one-out validation approach or a bootstrapping approach is typically used to test algorithm accuracy on data from participants that are held out from algorithm development (26,27,41). Nevertheless, these approaches do not ensure that datasets used for 'algorithm development' and 'algorithm testing' are completely independent. Both datasets share the commonality of activities. A recent study tested the accuracy of an activity type classification algorithm on an independent dataset (42). Freedson et al. (42) developed a neural network algorithm (nnet) using data collected at the University of Massachusetts (UMass) and applied it to data from different activities collected at the University of Tennessee (UTenn). The accuracy of the nnet for activity type recognition was 80.7% (42). When data from both UMass and UTenn were combined to develop an nnet algorithm and the leave-one-out validation was conducted, the accuracy increased to 97.3% (42). This result alone suggests that testing the algorithm on the same dataset used for its development leads to inflation of classification accuracy. Still, the study is a proof of concept that machine learning models can accurately detect activity type from a dataset containing different activities that share similar categorical characteristics (e.g. locomotion, household).

The reduction in accuracy observed by Freedson et al. (42) also preludes the probability of greater declines in classification accuracy of algorithms in free-living conditions given differences in both duration and performance of activities. In fact, past studies have shown that accuracy of algorithms developed on laboratory data decline substantially when applied to free-living conditions. Gyllensten and Bonomi (64) showed reduced accuracy of three machine learning models developed on laboratory data when applied to free-living accelerometer data. However, the reduction in accuracy reported by Gyllensten and Bonomi (64) was smaller (~16-20%) than the ~40-46% reduction we observed. Some factors may have contributed to a greater decline in accuracy seen in our study, including type of activity monitor, monitor placement, size of training dataset (theirs: >246 hours, ours: ~27 hours), and the criterion measure used. An important observation made by Gyllesten and Bonomi (64) was that activities in free-living conditions exhibit a higher degree of overlapping characteristics in their acceleration features when compared to activities performed in the lab. This may partially explain why locomotion in our study was markedly misclassified as household activity in the free-living condition.

A possible solution to address this problem is to use free-living accelerometer data to train machine learning algorithms for classification of activity type. This was highlighted in a previous study by Ermes et al. (63), where the authors reported substantial improvement (~17%) in the accuracy of their machine learning algorithms when free-living data were included in the training dataset. We employed this same approach and observed similar improvements in accuracy of our SVM and RF algorithms (9-14%). However, the improvement observed in our study was not adequate to consider our algorithms sufficiently accurate for use in free-living conditions. Our algorithms yielded high misclassification rates between household activities and locomotion, and also between standing and household activities. Perhaps, refinement of the direct observation (DO) system may yield more accurate free-living PA data that can be used to improve the training of our machine learning algorithms. Our current DO system does not allow for post observation coding, which is important for correcting miscoded data or for improving labeling of acceleration signals. The possibility of recoding activities using video records may be beneficial. With our system, very short activities (transitions) are often misclassified whereas with a video system even these short activities can be recoded more accurately, for example, transitions from sitting to standing. While accelerometer data corresponding to these transitions are easily matched in laboratory conditions, they are problematic when coding in real-time free-living conditions. If accelerometer data were matched with exact transition times, models that work with transition points rather than sliding windows could be developed from free-living data (32). This is central to improve accuracy of our algorithms for classifying activity type in shorter intervals. In this study, we attempted to classify activity using different interval durations (5 to 30 s). However, classification accuracy was reduced when using intervals of 5 or 10 s. With the recent advent of monitors that provide raw acceleration signals, studies in laboratory have been able to show that accurate classification of activity type can be done in intervals as short as 6.12 s (110). Considering the vast amount of data that can be collected with the current activity monitors (e.g. 100 Hz), it is realistic to classify activity type in short intervals during free-living conditions. Yet, much work is needed and our algorithms are far from being at that stage.

The results from the current study provide directions for future research developing machine learning algorithms to process accelerometer data. Random forest models might be a better option than support vector machine models for classifying activity type in free-living conditions. In addition, the results from this study reveal that monitor placement did not produce substantial differences in correct classification rates of the algorithms. Overall, the major inaccuracies of the algorithms were for standing and recreational activities (Table 5.4 a and b). These inaccuracies were slightly worse for algorithms using wrist accelerometer data. In free-living conditions, standing is usually accompanied of random body movements, which generates signal noise in the accelerometer data. Signal noise is usually greater for random arm movements (e.g. during conversation). This may explain the lowest correct classification rates for standing by algorithms using wrist accelerometer data. For recreational activities, it is likely that the algorithms are inaccurate in recognizing these activities because of the similarity of their acceleration signal with those from other activity categories. Recreational activities are usually a combination of standing, locomotion, and intermittent movement. Thus, it is not surprising that our algorithms did not present a systematic misclassification of recreational activities with another particular activity. In fact, table 5.5 shows that the misclassifications of recreational activities were dispersed among all other activity categories (standing, household, locomotion, and sedentary behavior). A possible solution for improving detection of standing and recreational activities may be to use hybrid models that first detect transition points and then apply sequential classification techniques, such as hidden Markov modeling to predict the most probable event to follow (32,59). Future studies will need to further examine the use of these approaches in improving activity type classification in older adults.

One strategy we used to increase classification accuracy was to reduce the number of activity categories from five to three categories (Table 5.4 and 5.6). Given the inability of the models in accurately classifying standing and recreational activity, we decided to combine standing with sedentary behavior, and recreational activity with household activity (Table 5.6). It is not ideal to combine sedentary behavior with standing posture because there are studies demonstrating independent effects of sedentary behavior on health outcomes (85–87). Also, objective assessment of both types of activities is instrumental in identifying older adults who present with fear of falling (141,142). This reinforces the importance of refining the algorithms in order to accurately classify the original five activity categories we proposed in laboratory conditions. This would allow researchers to employ our algorithms in studies investigating associations of types of activities with health and functional outcomes.

Although the linear mixed models analyses showed that predicted and actual time spent in different activity categories were not significantly different, caution is warranted as estimates were not precise overall (wide 95% CI) (Figures 5.3 and 5.4). In addition, percent correct classification rates of our free-living algorithms were predominantly modest (≥50% and <70%). In fact, the current accuracy level of our algorithms only allows us to recommend the use of our RF_{FL} algorithm for classification of three activity categories. Even so, different algorithms need to be used for capturing different activity categories, as none of the algorithms provide acceptable or high accuracy for all three categories. Table 5.8 provides a qualitative classification of accuracy of the RF_{FL} algorithms for each type of activity. Note that the only activity category that is consistently classified with an acceptable (≥70% and <80%) or high accuracy level (≥ 80%) by all the algorithms is the sed/stand category. For the ‘moving intermittently’ category, only the RF_{FL} hip algorithm using 20 s classification intervals provides estimates with acceptable accuracy in free-living older adults. Similarly, only the RF_{FL} hip algorithm using 30 s classification intervals provides acceptable accuracy for locomotion estimation. The ability of accurately estimating locomotion time is essential for using another one of our algorithms that predicts locomotion speed from accelerometer data (RF_{speed} algorithm).

In this respect, we used the RF_{FL} hip algorithm (3 activity groups) to obtain estimates of total time spent in free-living locomotion and then applied a previously developed RF algorithm to predict free-living locomotion speed (RF_{speed} hip). Predicted locomotion speed was correlated to speed in the 400 m walk. The results indicated that average free-living locomotion speed was weakly correlated to speed in the 400 m walk (r=0.22). On the other hand, highest free-living locomotion speed was moderately correlated with speed in the 400 m walk (r=0.55). It is important to note that only five participants in this study walked for at least 10 min. For these participants, ‘average free-living locomotion speed’ and ‘maximum free-living locomotion speed’ matched the pattern of lowest to highest speed in the 400 m walk, except for one participant. These results suggest a potential application of using activity type classification algorithms for identifying older adults with high mobility levels as well as those at higher risk for mobility disability. There is evidence that speed in the 400 m walk is associated with survival time and mortality in older adults (54). Thus it may be advantageous for future studies to use our locomotion speed prediction algorithm to examine associations of free-living locomotion speed with health, function and mortality in older adults.

This study has limitations. Our sample only included healthy and ambulatory older adults; therefore, our results are not generalizable to other populations. Participants were observed for a total of ~27 h and it is likely that more observation time and participants are needed for covering a broader range of activities that can possibly be performed in free-living conditions. Lastly, our current direct observation system does not allow for recoding activity type and duration, which is important for providing data of higher quality to train machine learning algorithms.

Summary

Our first hypothesis was that our machine learning algorithms developed in the laboratory would classify activity type in free-living older adults with similar accuracy as previous studies (~ 70% accuracy) (62,63). This hypothesis was not supported. The classification accuracy of the RF_{Lab} and SVM_{Lab} algorithms for activity type ranged from 49% to 55% in free-living older adults.

We also hypothesized that classification algorithms developed with free-living data would classify activity type in free-living older adults more accurately than algorithms developed in laboratory. This hypothesis was supported in this study. The RF_{FL} and SVM_{FL} algorithms for classification of five activity categories were more accurate in free-living older adults than the RF_{Lab} and SVM_{Lab} algorithms. However, only the RF_{FL} ankle and SVM_{FL} ankle algorithms achieved an overall correct classification rate of 70% when classifying five activity categories on a 30 s basis. Nevertheless, the RF_{FL}

algorithms reached up to 76% overall accuracy when classifying only three activity groups.

In this study, we also included an exploratory analysis. We investigated whether locomotion speed predicted by a machine learning algorithm was correlated to speed in the 400 m walk. Our analysis indicated a weak correlation between average locomotion speed predicted by the RF_{speed} algorithm and speed in the 400 m walk. However, the correlation was moderate for maximum locomotion speed predicted by the RF_{speed} algorithm and speed in the 400 m walk. It is important to note that only 5 participants walked for at least 10 min during the direct observation period. Therefore, it is reasonable to state that our results are promising and that future studies should further investigate the potential of using machine learning algorithms to predict locomotion speed in free-living conditions.

In conclusion, the results from this study suggest that our algorithms are currently not sufficiently accurate for assessment of free-living PA in older adults. It is necessary to further improve their accuracy, which may be possible by implementing a modified direct observation system that allows for recoding of data. This may allow for applying algorithms that detect point of transition to improve differentiation of the start and end of activities, and thus, minimizing confusion by the algorithms. Once refined and operational, activity type classification algorithms may have implications for assessing activity characteristics in older adults (e.g. locomotion speed) that are typically assessed using physical performance tests. Based on the current study, we recommend that future studies develop activity type classification algorithms using free-living accelerometer data and also employ a more sophisticated direct observation system as a criterion measure.

Tables

Table 5.1: Participant characteristics

Participant characteristics	
n	15 (9F, 6M)
Age (years)	70.0 ± 4.3
Body Mass (Kg)	74.5 ± 11.5
Height (cm)	169.8 ± 9.9
BMI ($Kg{\cdot}m^{-2}$)	26.0 ± 4.3
PA score (0 to 7)	4.6 ± 1.7
5 Chair stands – SPPB (s)	8.7 ± 1.2
8 foot walk – SPPB (s)	2.5 ± 0.4
400 m – walk speed ($m{\cdot}s^{-1}$)	1.2 ± 0.2

Values are mean and standard deviation. The score on the balance test from the SPPB is not provided in the table, as all participants were required to hold in each of the standing positions for ten seconds (max score).

Table 5.2: Activity menu from personal digital assistant (PDA) and their respective Activity categories

Activity (From PDA menu)	Activity Category*
• Standing • Standing with upper body movement	Standing
• Lying • Sitting • Sitting with upper body movement • Driving	Sedentary
• Moving intermittently • Household activities	Household
• Walking • Walking with a load • Walking incline • Stairs	Locomotion
• Aerobic Exercise • Resistance Exercise • Balance Exercise	Recreational

Left column exhibits list of activities from the PDA that observers used to code activity type in this study. Right column displays the activity category labels used to train machine learning algorithms from this study.

* Labels used to train new algorithms

Table 5.3: Performance of lab-based and free-living SVM and RF algorithms

			Lab- and Free-living Algorithms					
			Activity Groups (5 Classes)					
			Standing	**Sedentary**	**Household**	**Locomotion**	**Recreational**	**Overall**
20 second window	**Lab**	**SVM_{Lab} Hip**	0%	68%	71%	49%	13%	**49%**
		SVM_{Lab} Wrist	1%	73%	72%	33%	21%	**49%**
		SVM_{Lab} Ankle	0%	79%	87%	37%	20%	**55%**
	FL	**SVM_{FL} Hip**	38%	82%	63%	76%	24%	**64%**
		SVM_{FL} Wrist	10%	75%	73%	65%	20%	**58%**
		SVM_{FL} Ankle	52%	87%	65%	72%	41%	**69%**
	Lab	**RF_{FL} Hip**	0%	62%	82%	52%	17%	**51%**
		RF_{FL} Wrist	1%	71%	73%	34%	26%	**49%**
		RF_{FL} Ankle	0%	76%	87%	39%	19%	**54%**
	FL	**RF_{FL} Hip**	37%	81%	68%	80%	25%	**66%**
		RF_{FL} Wrist	11%	81%	70%	74%	22%	**61%**
		RF_{FL} Ankle	51%	84%	64%	69%	39%	**66%**

Values are percent of total time correctly identified for each activity group across all participants. Overall percent correct classification (last column on the right side) is percent correct classification across all activities and participants. Lab: laboratory, FL: free-living. First two clusters of rows display recognition accuracy for support vector machine algorithms developed with laboratory accelerometer data (SVM_{Lab}) and free-living accelerometer data (SVM_{FL}). Last two clusters of rows display recognition accuracy for random forest algorithms developed with laboratory accelerometer data (RF_{Lab}) and free-living accelerometer data (RF_{FL}).

Table 5.4: Performance of the newly developed SVM_{FL} and RF_{FL} algorithms for classification of 5 different activity groups according to different window length

a

Window Length		Support Vector Machine – Activity Group (5 Classes)	Standing	Sedentary	Household	Locomotion	Recreational	Overall
	5 sec	SVM_{FL} Hip	26%	79%	59%	64%	19%	**56%**
		SVM_{FL} Wrist	9%	70%	68%	61%	13%	**53%**
		SVM_{FL} Ankle	43%	78%	59%	54%	39%	**59%**
	10 sec	SVM_{FL} Hip	32%	82%	59%	70%	23%	**61%**
		SVM_{FL} Wrist	10%	74%	71%	64%	19%	**56%**
		SVM_{FL} Ankle	48%	82%	63%	60%	38%	**63%**
	20 sec	SVM_{FL} Hip	38%	82%	63%	76%	24%	**64%**
		SVM_{FL} Wrist	10%	75%	73%	65%	20%	**58%**
		SVM_{FL} Ankle	52%	87%	65%	72%	41%	**69%**
	30 sec	SVM_{FL} Hip	45%	82%	66%	75%	22%	**65%**
		SVM_{FL} Wrist	10%	78%	75%	67%	21%	**59%**
		SVM_{FL} Ankle	53%	87%	70%	77%	40%	**71%**

b

Window Length		Random Forest – Activity Group (5 Classes)	Standing	Sedentary	Household	Locomotion	Recreational	Overall
	5 sec	RF_{FL} Hip	28%	72%	61%	71%	16%	**57%**
		RF_{FL} Wrist	12%	73%	67%	69%	16%	**56%**
		RF_{FL} Ankle	43%	75%	59%	53%	38%	**58%**
	10 sec	RF_{FL} Hip	36%	77%	65%	73%	17%	**62%**
		RF_{FL} Wrist	12%	77%	69%	71%	22%	**59%**
		RF_{FL} Ankle	46%	81%	61%	58%	40%	**62%**
	20 sec	RF_{FL} Hip	37%	81%	68%	81%	25%	**61%**
		RF_{FL} Wrist	22%	81%	70%	74%	22%	**66%**
		RF_{FL} Ankle	51%	84%	64%	70%	39%	**67%**
	30 sec	RF_{FL} Hip	40%	83%	71%	81%	25%	**68%**
		RF_{FL} Wrist	10%	84%	73%	76%	24%	**63%**
		RF_{FL} Ankle	50%	86%	68%	78%	39%	**70%**

a) Recognition accuracy for the support vector machine algorithms developed with free-living accelerometer data (SVM_{FL}), b) recognition accuracy for the random forest algorithms developed with free-living accelerometer data (RF_{FL}). Values are percent of total time correctly identified for each activity group across all participants. Overall percent correct classification (last column on the right side) is percent correct classification across all activities and participants. Each cluster of rows displays performance of the algorithms according to a different classification interval. Within each cluster, each row represents performance of an algorithm developed with accelerometer data from a different monitor placement.

Table 5.5: Confusion matrix, and sensitivity and specificity values for free-living RF_{FL} ankle algorithm (30 s classification interval)

		RF_{FL} Ankle Algorithm				
		Predicted				
		Recreational	**Household**	**Locomotion**	**Sedentary**	**Standing**
Actual	**Recreational**	52	22	20	24	16
	Household	6	271	33	29	61
	Locomotion	2	68	266	2	1
	Sedentary	14	18	2	406	34
	Standing	2	75	1	49	125

Overall accuracy: 70%
(95% CI: 68% - 72%)

	Recreational	**Household**	**Locomotion**	**Sedentary**	**Standing**
Sensitivity	59%	50%	86%	69%	28%
Specificity	93%	89%	94%	92%	85%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Table 5.6: Performance of newly developed RF_{FL} algorithms for classification of 3 different activity groups using different classification intervals

			Random Forest			
			Activity Groups (3 Classes)			
			Sed/Stand	Moving inter	Locomotion	Overall
Window Length	5 sec	RF_{FL} Hip	78%	61%	61%	**69%**
		RF_{FL} Wrist	73%	63%	65%	**66%**
		RF_{FL} Ankle	76%	61%	47%	**65%**
	10 sec	RF_{FL} Hip	81%	63%	68%	**72%**
		RF_{FL} Wrist	73%	63%	65%	**68%**
		RF_{FL} Ankle	80%	62%	49%	**68%**
	20 sec	RF_{FL} Hip	81%	76%	67%	**75%**
		RF_{FL} Wrist	74%	63%	70%	**69%**
		RF_{FL} Ankle	84%	64%	58%	**72%**
	30 sec	RF_{FL} Hip	83%	68%	76%	**76%**
		RF_{FL} Wrist	77%	64%	65%	**70%**
		RF_{FL} Ankle	85%	68%	68%	**76%**

Recognition accuracy for the random forest algorithms developed with free-living accelerometer data (RF_{FL}). Values are percent of total time correctly identified for each activity group across all participants. Overall percent correct classification (last column on the right side) is percent correct classification across all activities and participants. Each cluster of rows displays performance of the algorithms according to a different classification interval. Within each cluster, each row represents performance of an algorithm developed with accelerometer data from a different monitor placement.

Table 5.7: Confusion matrix, and sensitivity and specificity values for the RF_{FL} hip algorithm (30 s classification interval)

		RF_{FL} Hip Algorithm		
		Predicted		
		Moving inter	**Locomotion**	**Sed/Standing**
Actual	**Moving inter**	364	45	126
	Locomotion	78	258	2
	Sed/Standing	124	1	600

Overall accuracy: 76%
(95% CI: 75% - 78%)

	Moving Inter	Locomotion	Sed/Standing
Sensitivity	64%	85%	82%
Specificity	83%	94%	86%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Table 5.8 Qualitative Classification of the Accuracy of the RF_{FL} Algorithms

			Activities		
			Sed/Stand	Moving inter	Locomotion
Classification Interval	5 sec	RF_{FL} Hip	✓	±	±
		RF_{FL} Wrist	✓	±	±
		RF_{FL} Ankle	✓	±	↓
	10 sec	RF_{FL} Hip	↑	±	±
		RF_{FL} Wrist	✓	±	±
		RF_{FL} Ankle	↑	±	↓
	20 sec	RF_{FL} Hip	↑	✓	±
		RF_{FL} Wrist	✓	±	±
		RF_{FL} Ankle	↑	±	±
	30 sec	RF_{FL} Hip	↑	±	✓
		RF_{FL} Wrist	✓	±	±
		RF_{FL} Ankle	↑	±	±

↑ High accuracy (≥ 80%)
✓ Acceptable accuracy (≥70% and <80%)
± Modest accuracy (≥50% and <70%)
↓ Low accuracy (<50%)

Sed/stand: sedentary behavior and standing; moving inter: moving intermittently
RF_{FL} algorithm: Random forest algorithm developed using free-living accelerometer data

Figures

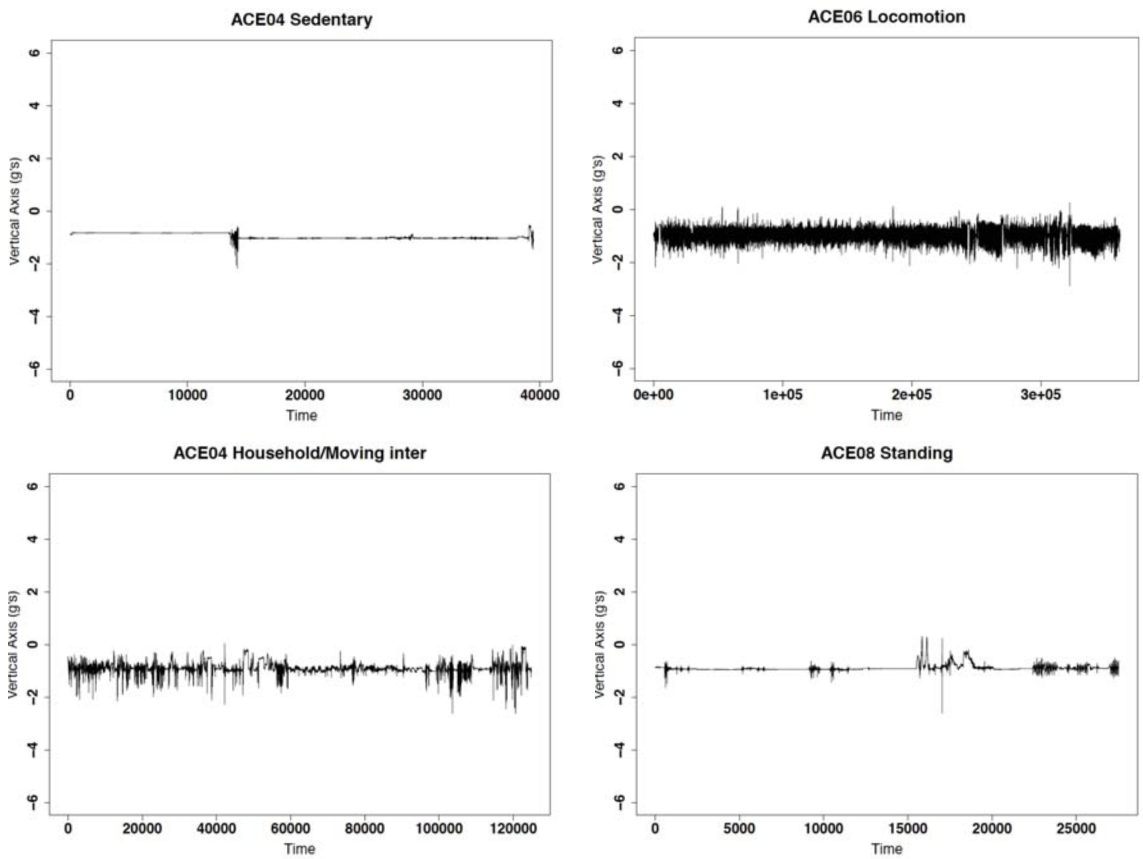


Figure 5.1: Free-living accelerometer signals with their direct observation labels

Signals are raw acceleration (g) from vertical axis collected at a sampling rate of 80 Hz. Each panel shows acceleration for a different activity (see panel title). *Y-axis* of each graph depicts acceleration and *x-axis* of each graph shows time duration for each activity (1/80 s). A single data point corresponds to 1/80th of a second.

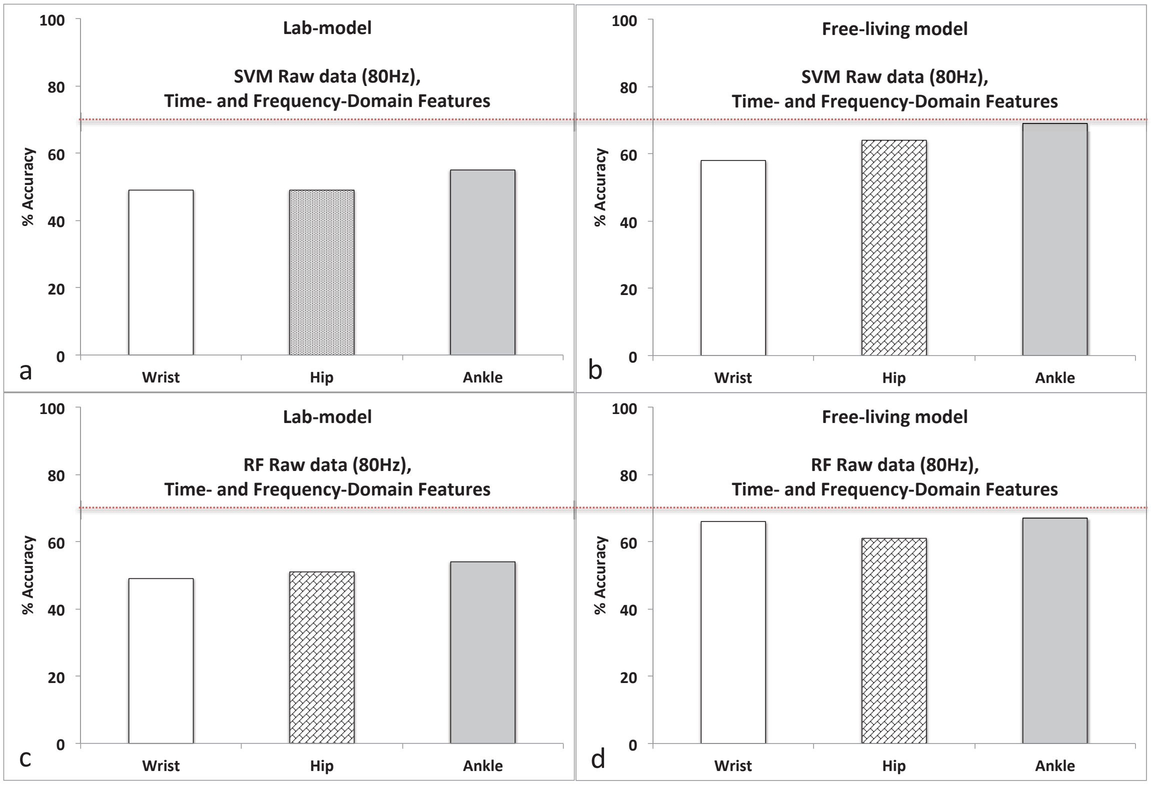


Figure 5.2: Performance of lab-based SVM and RF algorithms and free-living SVM and RF algorithms.

Left panels show performance of support vector machine (SVM) (panel a) and random forest (RF) (panel c) algorithms developed with laboratory accelerometer data. Right panels show performance of SVM (panel b) and RF (panel d) algorithms developed with free-living accelerometer data. The *y-axis* of each figure is overall percent correct classification of activities for combined data from all participants. The *x-axis* displays the bars for different monitor placement. The dotted line indicates 70% percent correct classification. This was the accuracy level (minimum) we aimed for in this study.

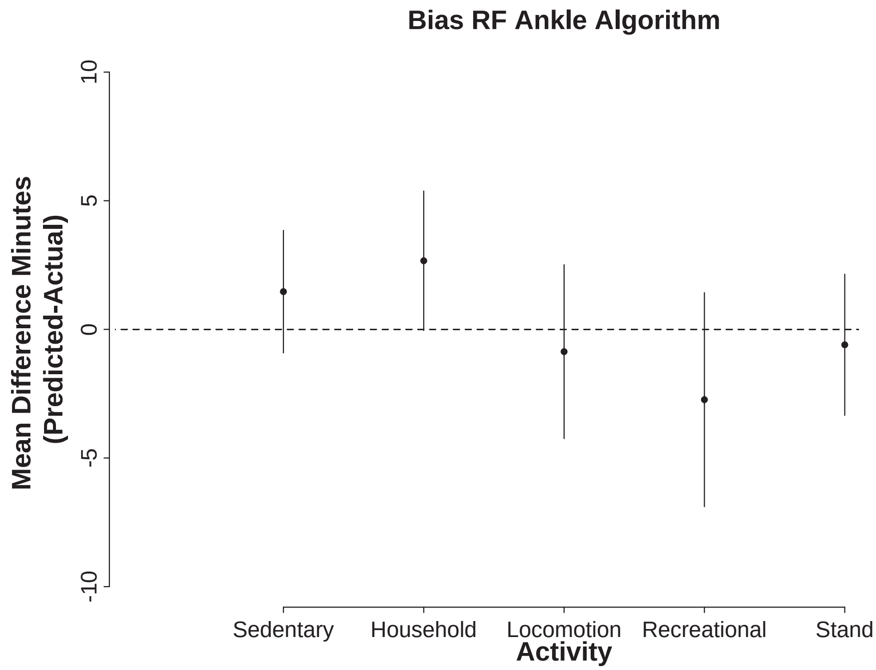


Figure 5.3: Bias of RF_{FL} ankle algorithm for the 5 activity groups

The *y-axis* displays mean difference in minutes (bias) between predicted minus actual time spent in different activity categories. The *x-axis* displays the different activity categories used in the current study. Black dots are mean values and error bars are 95% confidence intervals (CI). Linear mixed models indicated that estimates were not significantly different than zero. Observe that 95% CI cross zero for all activity categories ($p > 0.05$). Values are relative to 118 ± 19 min of direct observation (Sedentary: 33.6 ± 18.7 min, Household: 22.6 ± 12.2 min, Locomotion: 24.3 ± 30.7 min, Recreational: 9.4 ± 19.5 min). Private time was 3.8 ± 6.8 min.

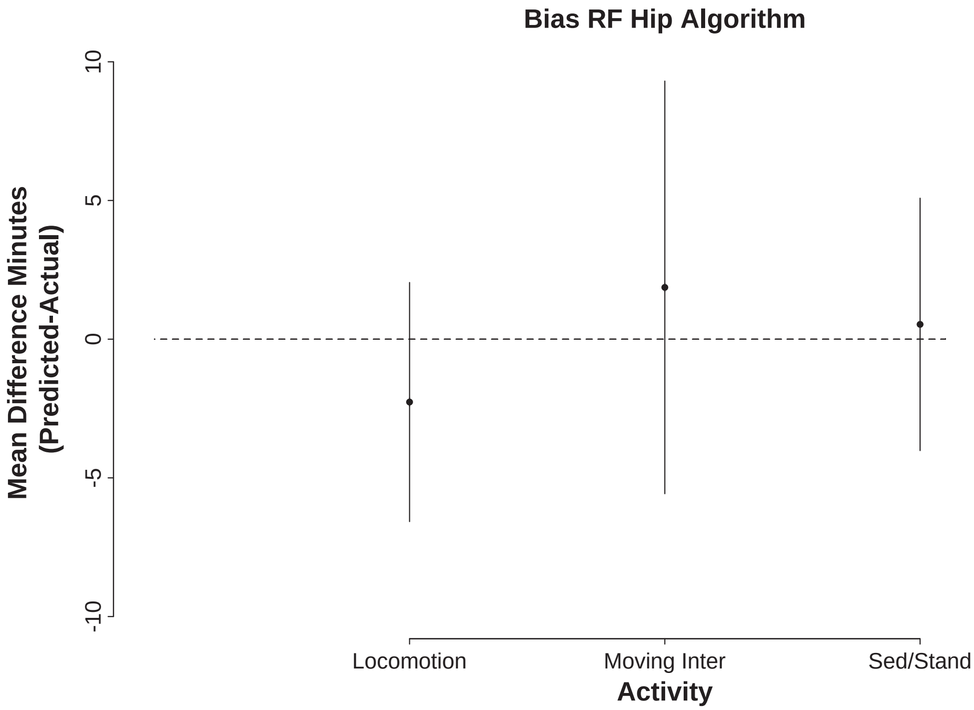


Figure 5.4: Bias of RF_{FL} hip algorithm for the 3 activity groups

The *y-axis* displays mean difference in minutes (bias) between predicted minus actual time spent in different activity categories. The *x-axis* displays the different activity categories used in the current study. Black dots are mean values and error bars are 95% confidence intervals (CI). Linear mixed models indicated that estimates were not significantly different than zero. Observe that 95% CI cross zero for all activity categories ($p > 0.05$). Values are relative to 118 ± 19 min of direct observation (Sed/Stand (sedentary and standing): 58.0 ± 26.4.7 min, Moving Inter (moving intermittently): 32.0 ± 21.7 min, and Locomotion: 24.3 ± 30.7 min). Private time was 3.8 ± 6.8 min.

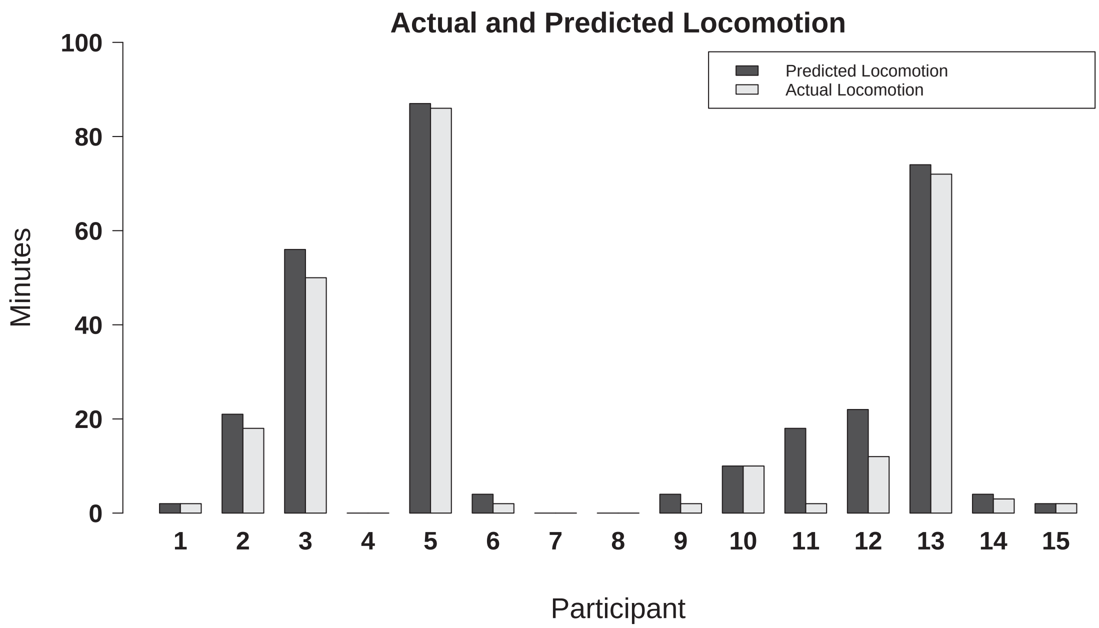
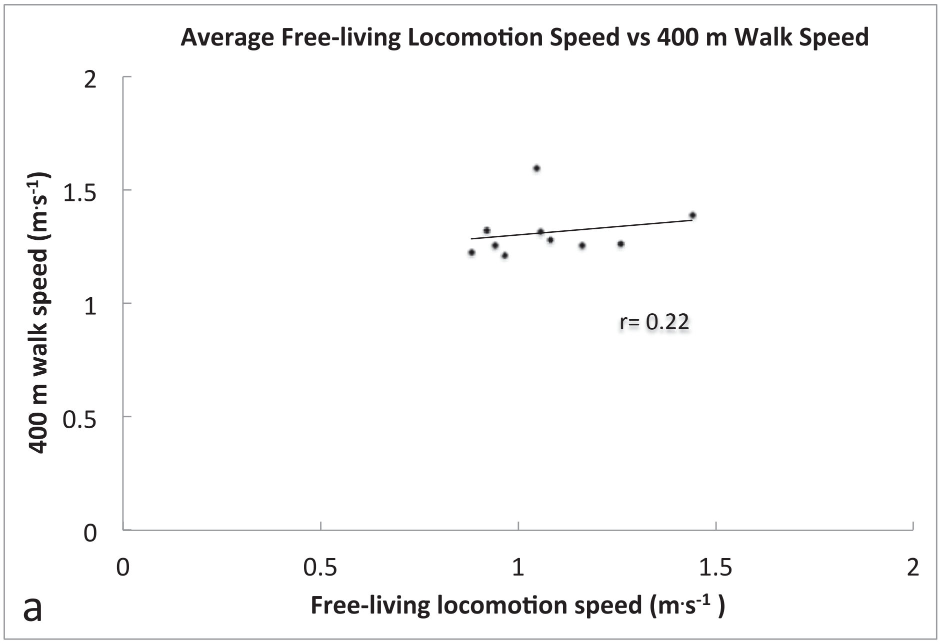


Figure 5.5: Time spent in locomotion for each individual according to RF_{FL} hip algorithm and Direct Observation (Actual)

Dark grey bars represent locomotion time predicted by RF_{FL} hip algorithm. Light grey bars represent locomotion time assessed by direct observation. Note that participants 11 and 12 drive the significant difference (t-test, $p<0.05$) between predicted and actual locomotion time. When these participants are removed from the analysis, difference is no longer significant.



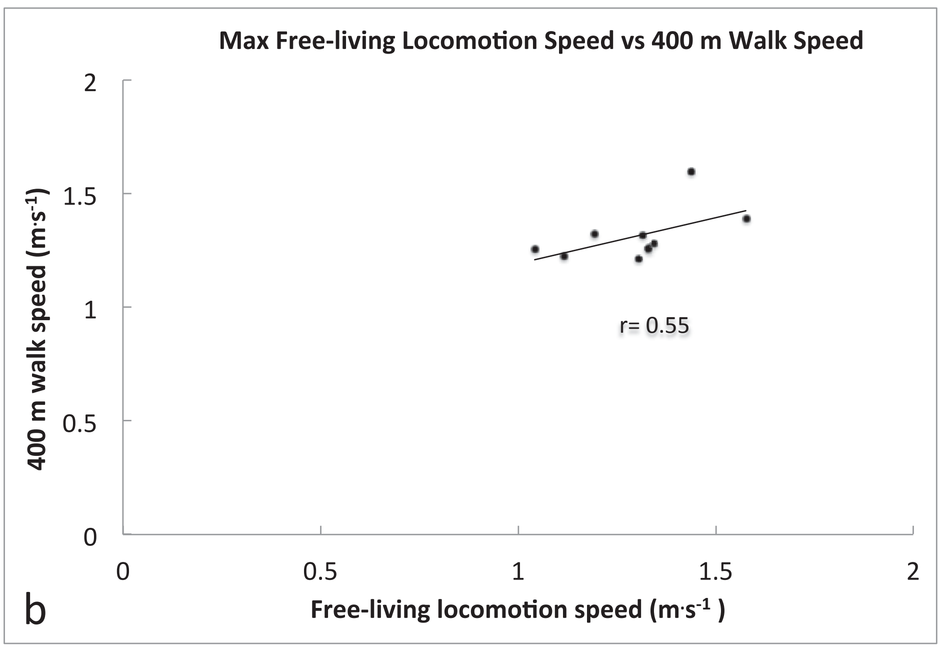


Figure 5.6: Correlations between free-living locomotion speed and 400 m walk speed.

a) Average free-living locomotion speed and 400 m walk speed, b) Maximum free-living locomotion speed and 400 m walk speed. Note: Most of the participants presented more than one bout of locomotion, and thus, more than one value for locomotion speed. We did not correlate all these values with speed in the 400 m walk. This would violate the independence between data points from the predictor variable, which is an assumption for running a Pearson product-moment correlation

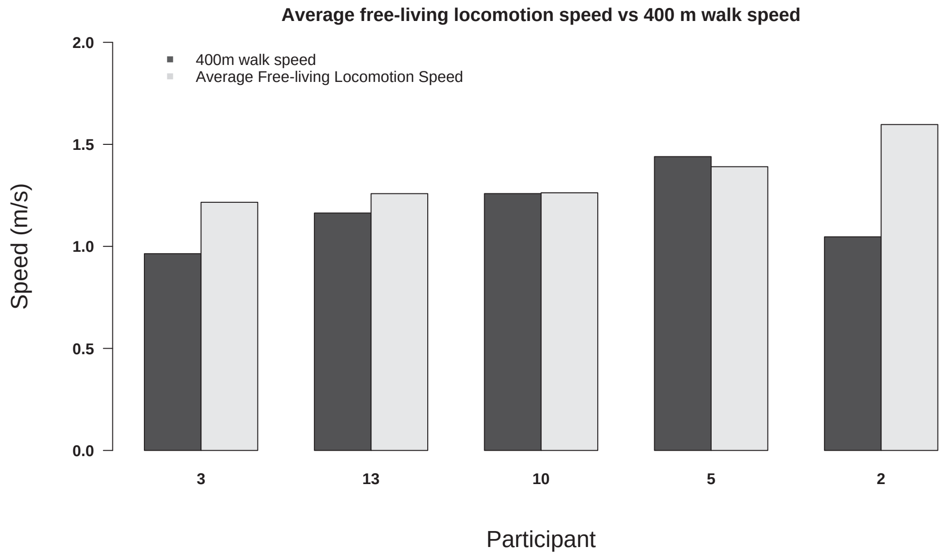


Figure 5.7 Free-living locomotion speed versus 400 m walk speed in participants performing locomotion bouts of at least 10 min.

Dark grey bars represent predicted locomotion speed. Light grey bars represent speed in the 400 m walk. *Top figure:* Average free-living locomotion speed versus 400 m walk speed. Average free-living locomotion speed was computed from all bouts of locomotion using the RF_{speed} hip algorithm developed in Study 1. *Bottom figure:* Maximum free-living locomotion speed versus 400m walk speed. Maximum free-living locomotion speed was the highest speed achieved by participants in free-living conditions. This was obtained among all locomotion bouts they performed. Note that all five participants performed bouts of locomotion lasting for at least 10 min. This cutoff point was selected based on the PA guidelines for Americans, in which the minimum duration of a meaningful PA bout is 10 min.

CHAPTER VI

SUMMARY

The main goal of this dissertation was to develop and test machine learning algorithms to process physical activity accelerometer data in older adults. In Study 1, we used a semi-structured activity protocol to develop and test machine learning algorithms to classify activity type and intensity from accelerometer data in older adults. In Study 2, the accuracy of the activity type classification algorithms was tested in free-living conditions and new algorithms were developed using free-living accelerometer data.

Study 1

In Study 1, it was hypothesized that our algorithms would accurately predict activity type in older adults during laboratory conditions. This hypothesis was supported. The machine learning algorithms predicted activity type with accuracy ranging from 87% to 96% when using both time- and frequency- domain features. Our data are in accordance with results from previous studies in younger adults, where algorithms accurately classified activity type from high-resolution accelerometer data (41,110,128).

It was also hypothesized that our machine learning algorithms would accurately predict activity intensity in older adults. The results showed that the algorithms were accurate for prediction of METs (small bias and RMSE), with no algorithm producing estimates significantly different than actual METs. For prediction of multiples of RMR ($Mult_{RMR}$), the algorithms were less accurate, and 4 of the 9 algorithms produced estimates significantly different than actual $Mult_{RMR}$. Based on these results, we can conclude that hypothesis 2 was supported for MET prediction but only partially supported for prediction of multiples of RMR. Similar to our results, previous studies have accurately predicted METs from accelerometer data using machine learning algorithms in young adults under controlled laboratory conditions (26,42).

Two exploratory analyses were also conducted in Study 1. The first exploratory analysis examined the best monitor placement for classification of activity type from accelerometer data in older adults. It was found that wrist accelerometer data provided the highest overall accuracies for activity type classification, with percent correct classification rates of up to 96%. The second exploratory aim was to examine the correlation of locomotion speed predicted by machine learning algorithms with speed during the 400 m walk (convergent validity). Our results revealed that the RF_{speed} hip and RF_{speed} ankle algorithms accurately predicted locomotion speed from accelerometer data. Conversely, the RF_{speed} wrist algorithm was not accurate for locomotion speed prediction. Thus, we suggest that studies intending to predict locomotion speed from accelerometer data should place activity monitors either on the hip or ankle.

Study 2

It was hypothesized that the lab-based algorithms would classify activity type from accelerometer data in free-living older adults with similar accuracy as previous studies (~ 70% accuracy) (62,63). This hypothesis was not supported. In Study 2, the algorithms developed in the laboratory performed poorly when classifying activity type in free-living older adults. The laboratory algorithms presented correct classification rates lower than 50% for most of the activity types, except for sedentary and household activities.

The second hypothesis was that algorithms developed using free-living accelerometer data would classify activity type more accurately than algorithms developed using laboratory accelerometer data. This hypothesis was supported. The accuracy of the free-living models was higher than the laboratory models, with some free-living models achieving up to 76% correct classification rate (overall). Compared to the laboratory algorithms, the free-living models showed improvements of up to 14% in classification accuracy for activity type. A previous study by Ermes et al. (63) reported similar findings for younger adults. They reported an improvement of approximately 17% in activity type classification when models were trained on both laboratory and free-living data compared to models only trained on laboratory data.

An exploratory analysis was conducted in Study 2. We examined the correlation of free-living locomotion speed - predicted by one of our machine learning algorithms (RF_{speed} hip algorithm) - with speed in the 400 m walk. The results showed a weak correlation between average locomotion speed predicted by the RF_{speed} hip algorithm and speed in the 400 m walk. However, there was a moderate correlation between maximum predicted speed and speed in the 400 m walk. It is important to note that only 5 participants walked for 10 minutes or more. It is possible that a stronger correlation would be found if we had obtained more data on locomotion.

Significance and Future Directions

The current investigation addressed the lack of activity type classification algorithms for processing accelerometer data from commercially available activity monitors in older adults. Our results demonstrated that machine learning algorithms accurately predict activity type in older adults under laboratory conditions. However, there is a decline in recognition accuracy when these algorithms are used in free-living conditions. It is necessary to refine our algorithms before they can be applied to real free-living PA assessment. A previous study predicted activity intensity in college-age students using a hybrid method that first identifies bouts of activity and inactivity and then applies an ANN to the identified bouts in order to predict METs (133). This approach produced substantially better predictions of METs compared to directly applying the ANN to the accelerometer data (133). Perhaps, a similar approach could be used in our case. It is possible that our algorithms can be improved by using a technique that identifies start and end of an activity rather than using a sliding window classification technique, as was used in the current study. In this case, a modified direct observation system will be necessary to better label accelerometer data for training machine learning algorithms. A possible approach for improving direct observation is to use video recording, which would allow for post-observation recoding of activities. Future studies should consider the limitations of this study in order to develop more accurate algorithms for classifying activity type in free-living older adults.

An important finding from this study is that the wrist monitor did not produce the highest accuracy in predicting activity type in free-living older adults. Overall, algorithms for the wrist monitor data were not able to detect standing, locomotion, and recreational activity. This should be taken into consideration by researchers when processing NHANES wrist accelerometer data. Before applying activity type classification algorithms to NHANES data, it is necessary to develop robust models to minimize misclassifications of certain activities. This involves training models with a large volume of free-living data and testing accuracy in diverse situations in the free-living environment.

While our machine learning algorithms developed in Study 2 were somewhat inaccurate, we were still able to demonstrate a potential application of these algorithms in free-living older adults. Locomotion speed predicted by the RF_{speed} algorithm was correlated to speed in the 400 m walk. In addition, the trend in predicted locomotion speed matched that of speed in the 400 m walk for 4 of the 5 participants. This result highlights the potential of using machine learning algorithms to identify mobility characteristics of free-living older adults. There is compelling evidence that locomotion speed in the 400 m walk is related to survival time and mortality (54). Machine learning algorithms may allow researchers to examine if locomotion speed in free-living conditions is also related to these outcomes.

In conclusion, the results from this dissertation suggest that 1) free-living accelerometer data are necessary for training activity type classification algorithms for application in real world settings, 2) refinements of the algorithms and more accurate criterion methods will be required to attain adequate levels of accuracy in free-living conditions, and 3) once operational, machine learning algorithms may have important applications in helping to understand the influence of free-living mobility characteristics on health and functional outcomes.

APPENDICES

APPENDIX A

CERTIFICATION OF HUMAN SUBJECTS APPROVAL

University of Massachusetts Amherst
108 Research Administration Bldg.
70 Butterfield Terrace
Amherst, MA 01003-9242

Research Compliance
Human Research Protection Office (HRPO)
Telephone: (413) 545-3428
FAX: (413) 577-1728

Certification of Human Subjects Approval

Date: December 16, 2011

To: Jeffer Eidi Sasaki, Kinesiology

Other Investigator: Dinesh John, Kinesiology, Patty Freedson, Kinesiology

From: Priscilla Clarkson, Chair, UMASS IRB

Protocol Title:DEVELOPMENT AND VALIDATION OF ACCELERATION-BASED ACTIVITY CLASSIFICATION ALGORITHMS FOR OLDER ADULTS: A MACHINE LEARNING APPROACH
Protocol ID: 2011-1154
Review Type:EXPEDITED - NEW
Paragraph ID: 4
Approval Date: 12/16/2011
Expiration Date:12/15/2012
OGCA #:

This study has been reviewed and approved by the University of Massachusetts Amherst IRB, Federal Wide Assurance # 00003909. Approval is granted with the understanding that investigator(s) are responsible for:

Modifications - All changes to the study (e.g. protocol, recruitment materials, consent form, additional key personnel), must be submitted for approval in e-protocol before instituting the changes. New personnel must have completed CITI training.

Consent forms - A copy of the approved, validated, consent form (with the IRB stamp) must be used to consent each subject. Investigators must retain copies of signed consent documents for six (6) years after close of the grant, or three (3) years if unfunded.

Adverse Event Reporting - Adverse events occurring in the course of the protocol must be reported in e-protocol as soon as possible, but no later than five (5) working days.

Continuing Review - Studies that received Full Board or Expedited approval must be reviewed three weeks prior to expiration, or six weeks for Full Board. Renewal Reports are submitted through e-protocol.

Completion Reports - Notify the IRB when your study is complete by submitting a Final Report Form in e-protocol.

Consent form (when applicable) will be stamped and sent in a separate e-mail. Use only IRB approved copies of the consent forms, questionnaires, letters, advertisements etc. in your research.

Please contact the Human Research Protection Office if you have any further questions. Best wishes for a successful project.

APPENDIX B

INFORMED CONSENT DOCUMENT – STUDY 1

Informed Consent Document
University of Massachusetts, Amherst
Dept. of Kinesiology

Development and Validation of Acceleration-Based Activity Classification Algorithms for Older Adults: A Machine Learning Approach

Phase 1: *Validation of Physical Activity Monitors in a Laboratory Setting*

Introduction

You are invited to participate in a research study conducted by the Physical Activity and Health Laboratory in the Department of Kinesiology at the University of Massachusetts Amherst. In this study, we will develop methods to assess physical activity using accelerometer-based physical activity monitors. These monitors are lightweight devices that can be secured on different parts of the body (e.g. waist, wrist, and ankle) and provide acceleration signals in response to movement. The purpose of this study is to use advanced statistical methods in order to identify activity type and intensity from these acceleration signals. The study will involve two visits to the laboratory: 1) Informed consent visit (current visit) and a 2) Physical activity visit.

Eligibility

To participate in this study, you must be between 65 and 80 years of age and in relatively good health, characterized by the absence of major cardiovascular, neurological, metabolic, bone or muscular disorders. Volunteers will be automatically excluded if they present any of the following conditions: 1) congenital heart disease 2) myocardial infarction or stroke in the past year, 3) congestive heart failure, 4) chronic obstructive pulmonary disease, 5) insulin-dependent diabetes mellitus, 6) Parkinson's disease, 7) Alzheimer's disease or any type of dementia, 8) muscular dystrophy, 9) active cancer treatment (e.g. radiotherapy, chemotherapy), 10) current use of certain medications that affect metabolism or cardiovascular and hemodynamic responses to exercise, and 11) mobility-impairment. Other conditions may also prevent volunteers to participate in the study, but those will be analyzed on an individual basis. At this point, you have been previously screened and are being considered for participation in the study. Please read this document carefully. It contains information about the study and the risks involved.

Research Procedures

Visit 1, Written Informed Consent (30 minutes)

You will report to the Physical Activity and Health Laboratory to review this informed consent document that was approved by the University of Massachusetts Institutional Review Board. In addition to the written details provided in this informed consent document, you will be given a verbal explanation of the study. You will have ample time to review this document and to ask any questions you may have. If you agree to participate, you will be requested to sign and date this document and a copy of this form will be provided for your records. You will then complete a health history questionnaire, a physical activity and health status questionnaire, a physical activity readiness questionnaire, and a physical function questionnaire. Once you complete them, your height and weight will be measured. Finally, we will send a blank medical clearance form to your physician along with a description of the study. Your physician will

Initials: ______

read it and decide whether or not to grant permission for your participation in the study. The form will be returned containing no private information about your health condition. Approval or disapproval will be communicated with no details. If you wish to know the reason as to why your physician disapproved your participation in the study, you should contact him or her directly as we will not have access to that information. If approval is granted by your physician, we will proceed with scheduling you for visit 2.

Visit 2 – Physical Activity Visit (120 minutes)

For this visit, you will be required to fast for 4h prior to coming to the laboratory as well as to refrain from exercise for a 12h-period. After your arrival to the laboratory, you will be asked to sit and rest quietly for 5 minutes, which will be followed by a resting heart rate and blood pressure measurement. If both resting heart rate and blood pressure are within normal values, we will proceed and measure your resting metabolic rate using the MedGem analyzer. The MedGem is a portable device that will provide an estimate of your resting energy expenditure based on the difference in the volume of oxygen between the air that you inhale and exhale. You will be asked to lie motionless on your back on a dormitory bed. After 10-15 minutes a nose clip will be placed on your nose and you will breathe into the mouthpiece of the MedGem analyzer for approximately 10 minutes. You will be able to breathe normally while using the MedGem. This equipment will provide a final measure that reflects the amount of daily energy that you spend in a resting state. Once the measurement is completed, we will provide you with a snack and a juice. You will also have time to drink water and use the restroom.

We will then proceed and fit you with the Viasys Oxycon Mobile (Carefusion, Yorba Linda, CA) system, which will measure your energy expenditure during the activities. The system is routinely used in studies measuring physical activity-related energy expenditure in laboratory settings. It weighs less than two pounds and it is placed in a harness that you will wear on your back or trunk. In addition, you will wear a headgear and a facemask that will be connected to the portable system allowing collection of expired air. You will be able to breathe normally with the facemask in place and time will be provided for you to become accustomed to breathing while wearing the facemask. Instructions on how to complete the activities will be provided as you become accustomed to breathing while wearing the Viasys Oxycon Mobile system and facemask.

You will also wear several activity monitors and a heart rate monitor and transmitter belt while you perform the activities. The activity monitors are small, about the size of a pager/beeper, and they do not inhibit motion or participation in any activity. The activity monitors will be worn on the hip, thigh, ankle, and wrist and will be fastened with elastic belts or non-allergenic medical adhesives. The heart rate transmitter (elastic belt) will be worn at the chest level and the monitor (wrist watch) will be worn on the wrist. These devices are routinely used in physical activity assessment studies. The investigators will make every effort to ensure that you are comfortable with the equipment and procedures and they will confirm that you are ready to proceed for they begin the testing session.

For the activity protocol, you will be asked to complete one of the two activity routines listed in tables 1 and 2. Each activity will be performed continuously for a 5-minute period. You will be given a 4-5 minute rest period between activities. In addition, you will be asked to perform three postural transitions: lying down to sitting, sitting to standing, and standing to sitting. You will remain still in each posture for 30sec and at the researcher signal you will perform the transition. At the end of the testing session, the activity monitors, heart rate transmitter and belt assembly and the Viasys Oxycon Mobile will be removed.

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 03/15/2012 **Protocol #:** 2011-1154
Valid Through: 12/15/2012
IRB Signature: Nancy C. Smith

Initials: ______

Table 1. Activity routine 1

Routine 1	**Description**
Crossword puzzles	A crossword puzzle book will be given to the participants who will choose a crossword to try and solve in five minutes.
Self-care (miscellaneous)	Participants will be instructed to perform several self-care tasks: 1) make the bed, 2) put shoes on, 3) prepare snacks and a drink 4) take shoes off and put them away. Routine is repeated for 5 minutes.
Organizing the room	Researchers will scatter different objects around the room and participants will be instructed to pick them up and organize by type.
Gardening	Participants will be given a small shovel and trowel to plant artificial flowers in an outdoor dirt patch.
Carrying groceries	Participants will walk on a 20 meter course carrying two plastic bags (e.g. one for each hand) containing bottled water totaling 1-5% of their body mass (0.5 to 5.0 pounds).
400m walk	Participants will be instructed to walk at a self-selected pace on a 20 meter course until covering 400 meters.
Tai-Chi or Recreational dance	Tai-Chi: Participants will reproduce the movements of a Tai-Chi instructional DVD. Recreational dance: A lesson from a ballroom dancing instructional DVD will be used to guide the activity.

Table 2. Activity routine 2

Routine 2	**Description**
Playing cards	Participants will play *Crazy Eight* against the researcher in the lab.
Laundry	A laundry basket containing several pieces of clothes will be placed on a table. Participants will be instructed to fold and then place the clothes on a pile.
Dusting	We will scatter paper confetti over 4 desktops. Participants will use a duster and a dustpan to sweep the confetti off the desktops.
Vacuuming	Round paper confetti will be scattered around the room (12 m^2) and participants will be instructed to use a vacuum cleaner to clean the carpet.
Slow walk (~1.8 mph)	Participants will walk on a treadmill at $0.8\ m \cdot s^{-1}$.
400m walk	Participants will be instructed to walk at a self-selected pace on a 20 meter course until covering 400 meters.
Simulated bowling	Bowling pins will be arranged on a gym court and participants will be instructed to play as they would on a bowling alley. The researcher will rearrange the pins every time participants throw the ball.

Risks

There are minimal risks arising from participating in this study. The risks are the same encountered in any self-paced physical activity, which include muscular discomfort, loss of balance, ankle sprains, and dizziness. However, these risks are small in relatively healthy participants.

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 03/15/2012 **Protocol #:** 2011-1154
Valid Through: 12/15/2012
IRB Signature:

Initials: ______

Benefits
Participation in this study will provide no specific benefits to you. The results obtained from this study will benefit the research community by creating new methods to assess physical activity in older adults. These methods will allow for future investigations relating physical activity to important aspect of late-life, such as physical disability prevention.

Withdrawal
Even if you sign this document, you are free to withdraw your consent and no longer participate in the study at any time. Withdrawing from this study will not influence your ability to participate in other studies at UMass.

Compensation
You are not being compensated for this informed consent visit. You will receive $30 for completing visit 2. Note that if you stop before the end of it, you will be compensated according to the percentage completed. For example, if you only complete half of visit 2, you will receive $15 (50% of $30).

Medical Treatment
The University of Massachusetts does not have a program for compensating subjects for injury or complications related to human subject research but in the unlikely event of injury resulting directly from participation in this study, investigators will assist you in every way to ensure that you get proper medical treatment. Medical treatment will be available to you through the University Health Services for a fee. Investigators will aid you in every way to see that you receive proper medical attention.

Enrollment/Length of Study
We expect to finish this study in approximately 10 months (December 2011 to September 2012). However, your participation in the study is only expected to last for 2 to 4h (in the laboratory).

Confidentiality
The information obtained from this study will be treated as privileged and confidential. It will not be released except upon your written consent. No personal identifying information will be used in the analysis or presentation of the data. You will be assigned a numerical ID number at the beginning of the study and all individual data will be identified by ID number only. Your name and ID number will be recorded at the beginning of the study and this information will be placed in a file cabinet that will be locked and only accessible to study researchers.

Request for Further Information
If you have any questions or concerns about being in this study you should contact Jeffer Sasaki by phone (413-545-1583) or email (pahealth@kin.umass.edu).

Review Board approval: The University of Massachusetts Institutional Review Board has approved this study. If you have any concerns about your rights as a participant in this study you may contact the Human Research Protection Office via email (humansubjects@ora.umass.edu); telephone (413-545-3428); or mail (Office of Research Affairs, 108 Research Administration Building, University of Massachusetts, 70 Butterfield Terrace, Amherst, MA 01003-9242).

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 03/15/2012 **Protocol #:** 2011-1154
Valid Through: 12/15/2012
IRB Signature: Nancy C. Scott

Initials: ______

Consent for Photograph

We are requesting to take photograph(s) of the participants during the research protocol for using in presentations (e.g. PowerPoint slides) and publications (e.g. scientific papers). You are not obligated to provide permission and your decision has no effect on whether you are or not eligible for the study. If you do provide consent to be photographed, you will select and approve the photograph(s) before giving researchers the permission to use it/them. Your consent is given with the condition that the researchers will edit the photograph(s) in order to prevent any facial identifiable feature. In addition, the photograph(s) will only be used by researchers from the Physical Activity and Health Laboratory at the University of Massachusetts Amherst. You will be able to request them to stop using the photograph(s) in future if you decide to; however, they will not be able to change publications and presentations preceding the request date. After reading this statement, please check one of the options below.

Yes □ I give consent to be photographed. My consent is given with the following restrictions (if any):

____
____
____

No □ I DO NOT wish to be photographed, but I can still participate in this study

PLEASE READ THE FOLLOWING STATEMENT AND SIGN BELOW IF YOU AGREE

I have had the chance to ask any questions I have about this study and my questions have been answered. I have read the information in this consent form and I voluntarily agree to be in the study. There are two copies of this form. I will keep one copy and return the other to the Physical Activity and Health Laboratory.

____ ________
Signature Date

Study Representative Statement

I have explained the purpose of the research, the study procedures, the possible risks and discomforts, the possible benefits, and have answered any questions to the best of my ability.

____
Study Representative Name (print or type)

____ ________________
Signature Date

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 03/15/2012 Protocol #: 2011-1154
Valid Through: 12/15/2012
IRB Signature: Nancy C. Swett

Initials: ______

APPENDIX C

PERSONAL HEALTH HISTORY QUESTIONNAIRE

Participant ID: ____ Interviewer:_____

Personal Health History

Name: ____________________ Age: ________ Date: ______________

Race: ____________________ Ethnicity: ( ) Hisp/Latino ( ) Non-hisp/Latino

Street Address: ____

City: _______________________State: ______ Zip code: _____________________

Home-phone:____________________ Cell-phone:________________

Work-phone:______________

E-mail:________________________________

Emergency Contact Name: ______________________ Phone: ______________________

1) *Has a physician ever told you that you have any of the following: (Check Yes or No)*

Yes	No		If yes, explain:
____	____	High Blood Pressure	________________________________
____	____	Diabetes	________________________________
____	____	Epilepsy	________________________________
____	____	Asthma	________________________________
____	____	Heart Disease	________________________________
____	____	Other	________________________________

2) *Any recent surgery*? (circle one) YES NO

If yes, please explain:

3) *Do you suffer from any osteomuscular disorder or physical impairment?* (circle one) YES NO
If yes, please explain:

4) *Do you suffer from any cardiopulmonary disease that affects your breathing or your ability of normally performing common daily activities?* (circle one)
YES NO

If yes, please explain:

5) *Do you suffer from any psychological disorder?* (circle one) YES NO

If yes, please explain:

6) Are you currently taking any medications? (circle one) YES NO

(include vitamins, herbal remedies, over-the-counter medicine, prescriptions medicine, etc.)

Medication	**Purpose**	**How Much**	**How Often**

APPENDIX D

PHYSICAL ACTIVITY READINESS QUESTIONNAIRE

PHYSICAL ACTIVITY READINESS QUESTIONNAIRE (PAR-Q)

Please read the following questions carefully and answer each one honestly: check YES or NO.

YES	NO	
☐	☐	1. Has your doctor ever said that you have a heart condition and that you should only do physical activity recommended by a doctor?
☐	☐	2. Do you feel pain in your chest when you do physical activity?
☐	☐	3. In the past month, have you had chest pain when you were not doing physical activity?
☐	☐	4. Do you lose your balance because of dizziness or do you ever lose consciousness?
☐	☐	5. Do you have a bone or joint problem that could be made worse by a change in your physical activity?
☐	☐	6. Is your doctor currently prescribing drugs (for example, water pills) for your blood pressure or heart condition?
☐	☐	7. Do you know of any other reason why you should not do physical activity?

PAR-Q (Thomas, Reading, & Shephard, 1992)

APPENDIX E

PHYSICAL ACTIVITY STATUS QUESTIONNAIRE

Participant ID: ___________ **Date:** _______________

Physical Activity Status

Using the descriptions below, record the highest number (0 to 7) which best describes your general activity level during the **previous month.** If you did more than section 1, then move on to section 2, and so on. You want to pick the highest number in this list to represent your activity level.

Section 1: Did not participate regularly in programmed recreational sport or heavy physical activity.

0 Avoided walking or exertion, e.g. always used the elevator, drove whenever possible instead of walking.

1 Walked for pleasure, routinely used the stairs, occasionally exercised sufficiently to cause heavy breathing or perspiration.

Section 2: Participated regularly in recreation or work requiring modest physical activity, such as golf, horseback riding, calisthenics, gymnastics, table tennis, bowling, weight lifting, yard work.

2 10 to 60 minutes per week.

3 Over 1 hour per week.

Section 3: Participated regularly in heavy physical exercise such as running or jogging, swimming, cycling, rowing, skipping rope, running in place or engaged in vigorous aerobic activity type of exercise such as tennis, basketball, or handball.

4 Ran less than 1 mile per week or spent less than 30 minutes per week in comparable physical activity.

5 Ran 1 to 5 miles per week or spent 30 to 60 minutes per week in comparable physical activity.

6 Ran 5 to 10 miles per week or spent 1 to 3 hours per week in comparable physical activity.

7 Ran more than 10 miles per week or spent over 3 hours per week in comparable physical activity.

Physical Activity Status during the previous month (highest score): _____________________

APPENDIX F

PHYSICAL ACTIVITY SCALE FOR THE ELDERLY

PHYSICAL ACTIVITY SCALE FOR THE ELDERLY

Instructions: Please complete this questionnaire by either circling the correct response or filling in the blank. Here is an example:

During the past 7 days, how often have you seen the sun?

☐ Never ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

Answer all items as accurately as possible. All information is strictly confidential.

LEISURE TIME ACTIVITY

1. Over the past 7 days, how often did you participate in sitting activities such as reading, watching TV or doing handcrafts?

☐ Never ⟶ If yes, go to Question **#2** ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

1a. What were these activities?

____

1b. On average, how many hours per day did you engage in these sitting activities?

☐ Less than 1 hour ☐ 2-4 hours

☐ 1 but less than 2 hours ☐ More than 4 hours

2. Over the past 7 days, how often did you take a walk outside your home or yard for any reason? For example for fun or exercise, walking to work, walking the dog, etc.?

☐ Never ⟶ If yes, go to Question **#3** ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

2a. On average, how many hours per day did you spend walking?

☐ Less than 1 hour ☐ 2-4 hours

☐ 1 but less than 2 hours ☐ More than 4 hours

3. Over the past 7 days, how often did you engage in light sport or recreational activities such as bowling, golf with a cart, shuffleboard, fishing from a boat or pier or other similar activities?

☐ Never ⟶ If yes, go to Question **#4** ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

3a. What were these activities?

____

3b. On average, how many hours per day did you engage in these light sport or recreational activities?

☐ Less than 1 hour ☐ 2-4 hours

☐ 1 but less than 2 hours ☐ More than 4 hours

4. Over the past 7 days, how often did you engage in moderate sport and recreational activities such as doubles tennis, ballroom dancing, hunting, ice skating, golf without a cart, softball or other similar activities?

☐ Never ⟶ If yes, go to Question **#5** ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

4a. What were these activities?

____

4b. On average, how many hours per day did you engage in these moderate sport and recreational activities?

☐ Less than 1 hour ☐ 2-4 hours

☐ 1 but less than 2 hours ☐ More than 4 hours

5. Over the past 7 days, how often did you engage in strenuous sport and recreational activities such as jogging, swimming, cycling, singles tennis, aerobic dance, skiing (downhill or cross-country) or other similar activities?

☐ Never ⟶ If yes, go to Question **#6** ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

5a. What were these activities?

____

5b. On average, how many hours per day did you engage in these strenuous sport and recreational activities?

☐ Less than 1 hour ☐ 2-4 hours

☐ 1 but less than 2 hours ☐ More than 4 hours

6. Over the past 7 days, how often did you do any exercises specifically to increase muscle strength and endurance, such as lifting weights or pushups, etc.?

☐ Never ⟶ If yes, go to Question **#7** ☐ Sometimes (3-4 Days)

☐ Seldom (1-2 Days) ☐ Often (5-7 Days)

6a. What were these activities?

____

6b. On average, how many hours per day did you engage in exercises to increase muscle strength and endurance?

☐ Less than 1 hour ☐ 2-4 hours

☐ 1 but less than 2 hours ☐ More than 4 hours

HOUSEHOLD ACTIVITY

7. During the past 7 days, have you done any light housework, such as dusting or washing dishes?

☐ No ☐ Yes

8. During the past 7 days, have you done any heavy housework or chores, such as vacuuming, scrubbing floors, washing windows, or carrying wood?

☐ No ☐ Yes

9. During the past 7 days, did you engage in any of the following activities? Please answer YES or NO for each item.

9a. Home repairs like painting, wallpapering, electrical work, etc.

☐ No ☐ Yes

9b. Lawn work or yard care, including snow or leaf removal, wood chopping, etc.

☐ No ☐ Yes

9c. Outdoor gardening

☐ No ☐ Yes

9d. Caring for another person, such as children, dependent spouse, or another adult

☐ No ☐ Yes

WORK-RELATED ACTIVITY

10. During the past 7 days, did you work for pay or as a volunteer?

10a. How many hours per week did you work for pay and/or as a volunteer?

__________ Hours

10b. Which of the following categories best describes the amount of physical activity required on your job and/or volunteer work?

- ☐ Mainly sitting with slight arm movements. [**Examples:** office worker, watchmaker, seated assembly line worker, bus driver, etc.]

- ☐ Sitting or standing with some walking. [**Examples:** cashier, general office worker, light tool and machinery worker]

- ☐ Walking, with some handling of materials generally weighing less than 50 pounds. [**Examples:** mailman, waiter/waitress, construction worker, heavy tool and machinery worker.]

- ☐ Walking and heavy manual work often requiring handling of materials weighing over 50 pounds. [**Examples:** lumberjack, stone mason, farm or general laborer]

APPENDIX G

SF-36 HEALTH SURVEY

SF36 - HEALTH SURVEY

1. **In general, would you say your health is:**
    - Excellent
    - Very good
    - Good
    - Fair
    - Poor

2. **Compared to one year ago, how would you rate your health in general now?**
    - Much better now than one year ago
    - Somewhat better now than one year ago
    - About the same as one year ago
    - Somewhat worse than one year ago
    - Much worse than one year ago

3. **The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much? (Check one)**
    a. *Vigorous activities* (such as running, lifting heavy objects, participating in strenuous sports)
     Yes, Limited a lot Yes, Limited a little No, Not at all

    b. *Moderate activities* (such as moving a table, pushing a vacuum cleaner, bowling or playing golf)
     Yes, Limited a lot Yes, Limited a little No, Not at all

    c. Lifting or carrying groceries
     Yes, Limited a lot Yes, Limited a little No, Not at all

    d. Climbing several flights of stairs
     Yes, Limited a lot Yes, Limited a little No, Not at all

    e. Climbing one flight of stairs
     Yes, Limited a lot Yes, Limited a little No, Not at all

    f. Bending, kneeling, or stooping
     Yes, Limited a lot Yes, Limited a little No, Not at all

    g. Walking more than a mile
     Yes, Limited a lot Yes, Limited a little No, Not at all

    h. Walking several blocks
     Yes, Limited a lot Yes, Limited a little No, Not at all

i. Walking one block
 Yes, Limited a lot Yes, Limited a little No, Not at all

j. Bathing or dressing yourself
 Yes, Limited a lot Yes, Limited a little No, Not at all

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

a. Cut down on the amount of time you spent on work or other activities
 Yes No

b. Accomplished less than you would like
 Yes No

c. Were limited in the kind of work or other activities
 Yes No

d. Had difficulty performing the work or other activities (For example, took extra effort)
 Yes No

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your emotional problems (such as feeling depressed or anxious)?

a. Cut down on the amount of time you spent on work or other activities
 Yes No

b. Accomplished less than you would like
 Yes No

c. Didn't do work or other activities as carefully as usual
 Yes No

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

 Not at all
 Slightly
 Moderately
 Quite a bit
 Extremely

7. **How much bodily pain have you had during the past 4 weeks?**

 None Moderate
 Very mild Severe
 Mild Very severe

8. **During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?**

 Not at all Quite a bit
 A little bit Extremely
 Moderately

9. **These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks:**

a. Did you feel full of pep?

 All of the time Some of the time
 Most of the time A little of the time
 A good bit of the time None of the time

b. Have you been a very nervous person?

 All of the time Some of the time
 Most of the time A little of the time
 A good bit of the time None of the time

c. Have you felt so down in the dumps that nothing could cheer you up?

 All of the time Some of the time
 Most of the time A little of the time
 A good bit of the time None of the time

d. Have you felt calm and peaceful?

 All of the time Some of the time
 Most of the time A little bit of the time
 A good bit of the time None of the time

e. Did you have a lot of energy?

 All of the time Some of the time
 Most of the time A little bit of the time
 A good bit of the time None of the time

f. Have you felt downhearted and blue

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

g. Did you feel worn out?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

h. Have you been a happy person?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

i. Did you feel tired?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

10. During the past 4 weeks, how much of the time has your physical health or emotional health interfered with your social activities (like visiting with friends, relatives, etc.)?

- All of the time
- Most of the time
- A good bit of the time
- Some of the time
- A little of the time
- None of the time

11. How TRUE or FALSE is each of the following statements for you?

a. I seem to get sick a little easier that other people

☐ True ☐ False

b. I am as healthy as anybody I know

☐ True ☐ False

c. I expect my health to get worse

☐ True ☐ False

d. My health is excellent

☐ True ☐ False

APPENDIX H

MEDICAL CLEARANCE FORM

Medical Clearance Form

Dear Dr.

Under the supervision of Dr. Patty Freedson from the Department of Kinesiology at University of Massachusetts Amherst, I am conducting a study to develop new methods to objectively assess physical activity in older adults. All prospective participants are asked to complete and sign a Physical Activity Readiness Questionnaire (PAR-Q), a personal health history questionnaire and an informed consent document. After determining initial eligibility, we are requesting that each participant obtain their physician's clearance in order to perform the research protocol.

The research protocol will require participants to visit the laboratory to perform an activity routine (routine 1 or 2) while wearing activity monitors and a portable indirect calorimetry system. Most of the activities will be performed for 5 minutes at a self-selected pace and a 4-minute rest will be allowed after completion of each activity. Participants will be able to stop the protocol at any time and may also withdrawal from the study at any time if they choose so. Please read the list of the activities below and provide a decision of whether or not you approve the participation of your patient in our study.

Activity routine:

Routine 1	Routine 2
Crossword puzzles	Playing cards
Self-care (miscellaneous)	Laundry
Organizing the room	Dusting
Gardening	Vacuuming
Carrying groceries	Slow walk (~1.8 mph)
400m walk	400m walk
Tai-Chi or Recreational dance	Simulated bowling

*See page 3 for description of the activities.

If appropriate, we ask that you provide clearance for this individual for entry into this study. If you have any further questions, please contact Jeffer Eidi Sasaki at (413) 545-1583.

As a result of my examination of ___________________________________
(Participant's Name)

[ ] I approve his/her participation in the study
[ ] I do not approve his/her participation in the study

Comments:

____

____

____

______________________________________ ___________________
(Physician's Signature) (Date)

I _________________________________ give permission to my physician to approve/disapprove my participation in this study.

After completing this form please fax a copy to (413) 545-2906 or mail to Attn: Jeffer Eidi Sasaki, Dept. of Kinesiology, 30 Eastman Lane 110 Totman Building Amherst, MA 01003-9258

Activity description:

Activity	Description
Crossword puzzles	A crossword puzzle book will be given to the participants who will choose a crossword to try and solve in five minutes.
Playing cards	Participants will play *Crazy Eight* against the researcher in the lab.
Self-care (miscellaneous):	Participants will be instructed to perform several self-care tasks: 1) make the bed, 2) put shoes on, 3) prepare snacks and a drink 4) take shoes off and put them away. Routine is repeated for 5 minutes.
Laundry	A laundry basket containing several pieces of clothes will be placed on a table. Participants will be instructed to fold and then place the clothes on a pile.
Organizing the room	Researchers will scatter different objects around the room and participants will be instructed to pick them up and organize by type.
Dusting	We will scatter paper confetti over 4 desktops. Participants will use a duster and a dustpan to sweep the confetti off the desktops.
Gardening	Participants will be given a small shovel and trowel to plant artificial flowers in an outdoor dirt patch.
Vacuuming	Round paper confetti will be scattered around the room (12 m^2) and participants will be instructed to use a vacuum cleaner to clean the carpet.
Carrying groceries	Participants will walk on a 20 meter course carrying two plastic bags (e.g. one for each hand) containing bottled water totaling 1-5% of their body mass (0.5 to 5.0 pounds).
Slow walk	Participants will walk on a treadmill at 0.8 $m \cdot s^{-1}$.
Recreational dance	A lesson from a ballroom dancing instructional DVD will be used to guide the activity.
400m walk	Participants will be instructed to walk at a self-selected pace on a 20 meter course until covering 400 meters.
Tai-Chi	Participants will reproduce the movements

	of a Tai-Chi instructional DVD.
Simulated bowling	Bowling pins will be arranged on a gym court and participants will be instructed to play as they would on a bowling alley. The researcher will rearrange the pins every time participants throw the ball.

APPENDIX I

DESCRIPTION OF ACTIVITIES

Description of activities

Activity	Description
Crossword puzzles	A crossword puzzle book will be given to the participants who will choose a crossword to try and solve in five minutes.
Playing cards	Participants will play *Crazy Eight* against the researcher in the lab.
Self-care (miscellaneous):	Participants will be instructed to perform several self-care tasks: 1) make the bed, 2) put shoes on, 3) prepare snacks and a drink 4) take shoes off and put them away. Routine is repeated for 5 minutes.
Laundry	A laundry basket containing several pieces of clothes will be placed on a table. Participants will be instructed to fold and then place the clothes on a pile.
Organizing the room	Researchers will scatter different objects around the room and participants will be instructed to pick them up and organize by type.
Dusting	We will scatter paper confetti over 4 desktops. Participants will use a duster and a dustpan to sweep the confetti off the desktops.
Gardening	Participants will be given a small shovel and trowel to plant artificial flowers in an outdoor dirt patch.
Vacuuming	Round paper confetti will be scattered around the room (12 m^2) and participants will be instructed to use a vacuum cleaner to clean the carpet.
Carrying groceries	Participants will walk on a 20 meter course carrying two plastic bags (e.g. one for each hand) containing bottled water totaling 1-5% of their body mass (0.5 to 5.0 pounds).
Slow walk	Participants will walk on a treadmill at 0.8 $m \cdot s^{-1}$.
Recreational dance	A lesson from a ballroom dancing instructional DVD will be used to guide the activity.
400m walk	Participants will be instructed to walk at a self-selected pace on a 20 meter course until covering 400 meters.
Tai-Chi	Participants will reproduce the movements

	of a Tai-Chi instructional DVD.
Simulated bowling	Bowling pins will be arranged on a gym court and participants will be instructed to play as they would on a bowling alley. The researcher will rearrange the pins every time participants throw the ball.

APPENDIX J

BORG SCALE

Borg Scale

6	
7	**Very, very light exertion**
8	
9	**Very light exertion**
10	
11	**Fairly light exertion**
12	
13	**Somewhat hard exertion**
14	
15	**Hard exertion**
16	
17	**Very hard exertion**
18	
19	**Very, very hard exertion**
20	

APPENDIX K

INFORMED CONSENT DOCUMENT – STUDY 2

Informed Consent Document
University of Massachusetts, Amherst
Dept. of Kinesiology

Development and Validation of Acceleration-Based Activity Classification Algorithms for Older Adults: A Machine Learning Approach

Phase 2: Validation of Physical Activity Monitors in Free-living conditions

Introduction
You are invited to participate in phase 2 of the study entitled *"Development and Validation of Acceleration-Based Activity Classification Algorithms for Older Adults: A Machine Learning Approach"*. In phase 1, we were interested in developing advanced methods to analyze acceleration data collected with physical activity monitors in simulated free-living activities. In phase 2, we will test how well these methods assess physical activity in free-living conditions. This will require you to use the same lightweight activity monitors for a 7-day period. In one of those days, we will directly observe you for a 3h time-block. Please read this document carefully, it contains information about the study and the risks involved.

Eligibility
To participate in this phase of the study, you must meet the following conditions:
1) You participated in phase 1 of the study.
2) Your health status did not change since then.
3) You are willing to use physical activity monitors for seven days.
4) You are willing to be directly observed for a 3h time-block.

Research Procedures
Visit 1: Written informed consent (30 minutes)
You will report to the Physical Activity and Health Laboratory to review this informed consent document that was approved by the University of Massachusetts Institutional Review Board. In addition to the written details provided in this informed consent document, you will be given a verbal explanation of the study. You will have ample time to review this document and to ask any questions you may have. If you agree to participate, you will be requested to sign and date this document and a copy of this form will be provided for your records.

Wearing Physical Activity Monitors in Free Living Conditions (7 days)
You will be asked to wear physical activity monitors during a 7-day period. The monitors will be worn on the dominant hip and thigh. We will instruct you on how to wear these monitors. In addition, we will ask you to record (on a log) the times when you start and stop wearing the monitors as well as the times when you remove the monitors (e.g. shower, swimming). At the end of the 7-day period, you will return the monitors along with the monitor log to the Physical Activity and Health Laboratory.

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 12/16/2011 **Protocol #:** 2011-1154
Valid Through: 12/15/2012
IRB Signature:

Initials: ______

Direct Observation (180 min)
In one of the days that you will be wearing the monitors, we will schedule a 3h time-block to direct observe you in your free-living environment. During this period, you will wear two additional activity

monitors: one on your dominant wrist and another one on your dominant ankle. They will be removed once the 3h direct observation period is over. During direct observation, a trained observer will record your activities while keeping a reasonable distance and remaining inconspicuous. You will be able to request privacy at any time. In this case, the observer will wait until you are ready to resume your normal routine. If you feel uncomfortable about being observed, you may request data collection to be stopped at any time.

Risks
Participation in this study will not create any additional risks to those you could encounter in your everyday life, which may include loss of balance, fatigue, ankle sprains and dizziness. However, if any incident occurs to you while participating in this study, we will assist you to ensure that you get proper medical treatment at your own expense (see *"Medical Treatment").*

Benefits
Participation in this study will provide no specific benefits to you. The results obtained from this study will benefit the research community by creating new methods to assess physical activity in older adults. These methods will allow for future investigations relating physical activity to important aspect of late-life, such as physical disability prevention.

Withdrawal
Even if you sign this document, you are free to withdraw your consent and no longer participate in the study at any time. Withdrawing from this study will not influence your ability to participate in other studies at UMass.

Compensation
You will receive $20 for wearing the activity monitors during the 7-day period and for being directly observed for the 3h time-block. If you quit before completing the protocol, you will be compensated according to the percentage completed. For example, if you only complete half of the protocol, you will receive $10 (50% of $20). Be aware that you are not being compensated for this informed consent visit.

Medical Treatment
The University of Massachusetts does not have a program for compensating subjects for injury or complications related to human subject research but in the unlikely event of injury while participating in this study, investigators will assist you in every way to ensure that you get proper medical treatment at your expense.

Enrollment/Length of Study
This study is expected to have a total duration of approximately 10 months (December 2011 to September 2012). However, your participation in the study will only require a total commitment of approximately 8-days (in your free-living time).

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 12/16/2011 Protocol #: 2011-1154
Valid Through: 12/15/2012
IRB Signature: Nancy C. Swett

Initials: ______

Confidentiality
The information obtained from this study will be treated as privileged and confidential. It will not be released except upon your written consent. No personal identifying information will be used in the analysis or presentation of the data. You will be assigned a numerical ID number at the beginning of the study and all individual data will be identified by ID number only. Your name and ID number will be recorded at the beginning of the study and this information will be placed in a file cabinet that will be locked and only accessible to study researchers.

Request for Further Information
If you have any questions or concerns about being in this study you should contact Jeffer Sasaki by phone (413-545-1583) or email (pahealth@kin.umass.edu).

Review Board approval: The University of Massachusetts Institutional Review Board has approved this study. If you have any concerns about your rights as a participant in this study you may contact the Human Research Protection Office via email (humansubjects@ora.umass.edu); telephone (413-545-3428); or mail (Office of Research Affairs, 108 Research Administration Building, University of Massachusetts, 70 Butterfield Terrace, Amherst, MA 01003-9242).

PLEASE READ THE FOLLOWING STATEMENT AND SIGN BELOW IF YOU AGREE

I have had the chance to ask any questions I have about this study and my questions have been answered. I have read the information in this consent form and I voluntarily agree to be in the study. There are two copies of this form. I will keep one copy and return the other to the Physical Activity and Health Laboratory.

____ ________
Signature Date

Study Representative Statement
I have explained the purpose of the research, the study procedures, the possible risks and discomforts, the possible benefits, and have answered any questions to the best of my ability.

____
Study Representative Name (print or type)

____ __________________
Signature Date

University of Massachusetts Amherst-IRB
(413) 545-3428
Approval Date: 12/16/2011 **Protocol #:** 2011-1154
Valid Through: 12/15/2012
IRB Signature: Nancy C. Sweet

Initials: ______

APPENDIX L

SUPPLEMENTAL TABLES FOR LABORATORY ALGORITHMS

Confusion matrix, and sensitivity and specificity values for the laboratory ANN hip algorithm

		ANN Hip Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	325	0	5	1	0
	Sedentary	1	131	10	7	1
	Household	2	8	424	46	0
	Recreational	0	7	48	102	3
	Standing	0	4	0	2	5

Overall accuracy: 87%

(95% CI: 85%-88%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	89%	85%	65%	31%
Specificity	99%	98%	91%	94%	99%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory ANN Wrist algorithm

		ANN Wrist Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	322	0	8	1	0
	Sedentary	0	166	6	2	2
	Household	7	3	450	19	1
	Recreational	3	2	8	145	2
	Standing	0	1	0	1	8

Overall accuracy: 94%

(95% CI: 93% - 95%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	97%	96%	96%	86%	55%
Specificity	99%	99%	96%	98%	100%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory ANN Ankle algorithm

		ANN Ankle Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	305	0	6	1	0
	Sedentary	0	144	12	6	7
	Household	2	10	403	35	1
	Recreational	1	4	41	105	0
	Standing	0	4	1	0	5

Overall accuracy: 88%

(95% CI: 86% - 89%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	90%	96%	70%	30%
Specificity	99%	97%	96%	95%	99%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory RF Hip algorithm

		RF Hip Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	327	0	2	2	0
	Sedentary	0	137	6	6	0
	Household	2	9	438	31	0
	Recreational	0	6	69	84	0
	Standing	0	10	0	1	0

Overall accuracy: 87%

(95% CI: 86%-88%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	87%	85%	68%	0%
Specificity	99%	99%	93%	93%	99%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory RF Wrist algorithm

		RF Wrist Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	319	1	11	1	0
	Sedentary	0	165	6	3	3
	Household	3	3	456	17	1
	Recreational	0	2	10	147	1
	Standing	0	0	0	2	9

Overall accuracy: 94%

(95% CI: 93%-95%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	96%	94%	87%	51%
Specificity	99%	99%	97%	99%	100%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory RF Ankle algorithm

		RF Ankle Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	309	0	3	0	0
	Sedentary	0	152	11	5	2
	Household	3	5	417	26	1
	Recreational	2	0	56	92	0
	Standing	0	5	1	0	4

Overall accuracy: 89%

(95% CI: 88%-90%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	94%	85%	76%	52%
Specificity	99%	98%	94%	94%	99%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory SVM Hip algorithm

		SVM Hip Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	326	0	4	1	0
	Sedentary	0	137	5	7	0
	Household	1	13	436	30	0
	Recreational	0	13	59	88	0
	Standing	0	9	0	1	0

Overall accuracy: 88%

(95% CI: 86%-89%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	82%	86%	69%	0%
Specificity	100%	99%	93%	93%	99%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

Confusion matrix, and sensitivity and specificity values for the laboratory SVM Hip algorithm

		SVM Ankle Algorithm				
		Predicted				
		Locomotion	Sedentary	Household	Recreational	Standing
Actual	Locomotion	309	0	3	0	0
	Sedentary	0	157	8	5	0
	Household	1	11	420	20	0
	Recreational	0	3	51	97	0
	Standing	0	6	2	0	2

Overall accuracy: 90%

(95% CI: 89%-91%)

	Locomotion	Sedentary	Household	Recreational	Standing
Sensitivity	99%	89%	87%	80%	78%
Specificity	99%	99%	95%	94%	99%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.

APPENDIX M

SUPPLEMENTAL TABLE AND FIGURE FOR FREE-LIVING SVM ANKLE ALGORITHM

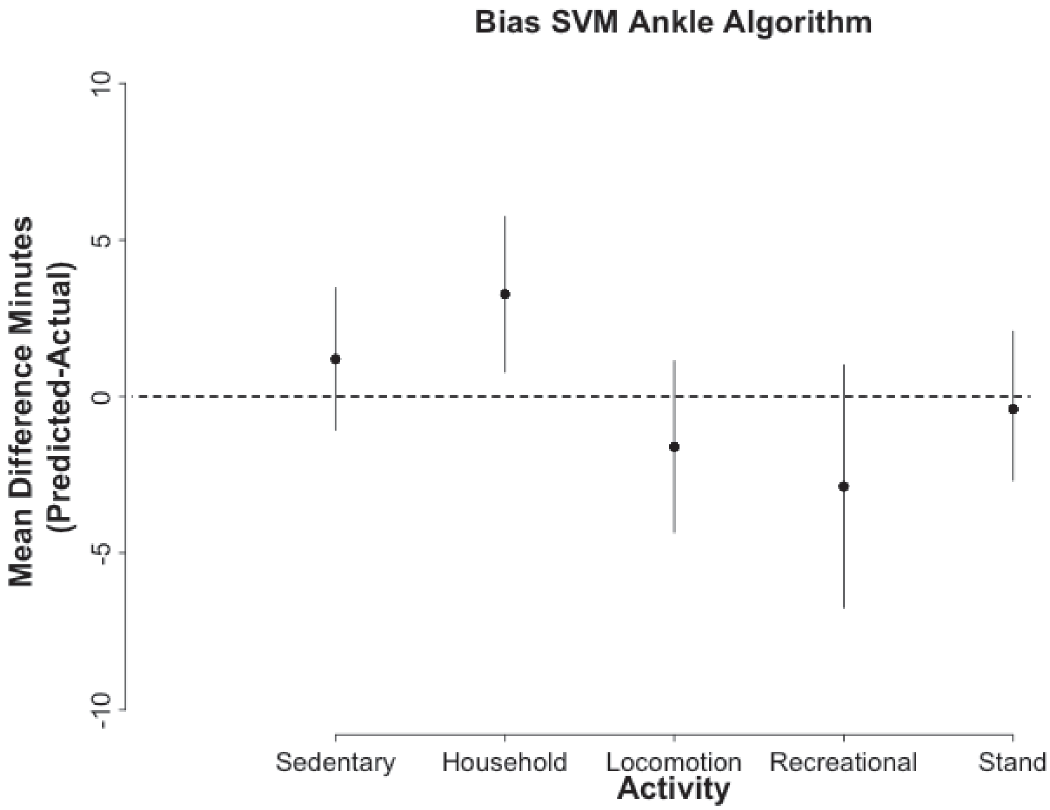
Confusion matrix, and sensitivity and specificity values for the free-living SVM ankle algorithm

		SVM_{FL} Ankle Algorithm				
		Predicted				
		Recreational	Household	Locomotion	Sedentary	Standing
Actual	Recreational	53	21	12	31	17
	Household	6	282	28	24	60
	Locomotion	2	73	260	1	2
	Sedentary	10	17	2	412	33
	Standing	2	81	0	34	134

Overall accuracy: 71%
(95% CI: 70% - 73%)

	Standing	Sedentary	Locomotion	Household	Recreational
Sensitivity	54%	82%	86%	59%	74%
Specificity	91%	94%	94%	89%	95%

Upper panel: Rows are actual activity and columns are predicted activity. Values are in minutes and combined for all participants. Shaded values are correctly classified minutes and values outside the diagonal line (shaded) are misclassified minutes. Middle panel: Overall accuracy indicates the percent correct classification of the algorithm for combined data of all activities and participants. 95% CI indicates the upper and lower bound of correct classification for 95% of the participants. Lower panel: Values are percent of detection by the algorithm. Note: Sensitivity identifies the number of true events that are correctly classified as such. Specificity identifies the number of false events that are correctly classified as false events.



Bias of the free-living SVM ankle algorithm for time spent in different activity groups

The *y-axis* displays mean difference in minutes (bias) between predicted minus actual time spent in different activity categories. The *x-axis* displays the different activity categories used in the current study. Black dots are mean values and error bars are 95% confidence intervals (CI). Linear mixed models indicated that estimates for household were significantly different than zero. Observe that 95% CI does not cross zero household activity ($p < 0.05$). All other estimates were not significantly different than actual time spent in the different activity categories. Values are relative to 118 ± 19 min of direct observation (Sedentary: 33.6 ± 18.7 min, Household: 22.6 ± 12.2 min, Locomotion: 24.3 ± 30.7 min, Recreational: 9.4 ± 19.5 min). Private time was 3.8 ± 6.8 min.

REFERENCES

1. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc. 2000 Sep;32(9 Suppl):S498–504.

2. Rolland YM, Cesari M, Miller ME, Penninx BW, Atkinson HH, Pahor M. Reliability of the 400-M Usual-Pace Walk Test as an Assessment of Mobility Limitation in Older Adults. J Am Geriatr Soc. 2004 Jun 1;52(6):972–6.

3. Verbrugge LM, Jette AM. The disablement process. Soc Sci Med. 1994 Jan;38(1):1–14.

4. Pope AM, Tarlov AR. Disability in America: Toward a National Agenda for Prevention. [Internet]. National Academy Press, 2201 Consitution Ave., N.W., Washington, DC; 1991 [cited 2011 Apr 28]. Available from: http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED336892

5. Buchner DM, Wagner EH. Preventing frail health. Clin Geriatr Med. 1992 Feb;8(1):1–17.

6. Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical Activity and Public Health in Older Adults. Med Sci Sports Exerc. 2007 Aug;39(8):1435–45.

7. Klein RJ, (US) NC for HS, Control C for D, (US) P. Healthy People 2010 criteria for data suppression. US Dept of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics; 2002.

8. Pandya A, Gaziano TA, Weinstein MC, Cutler D. More Americans Living Longer With Cardiovascular Disease Will Increase Costs While Lowering Quality Of Life. Health Aff. 2013 Oct 1;32(10):1706–14.

9. Zubritsky C, Abbott KM, Hirschman KB, Bowles KH, Foust JB, Naylor MD. Health-related Quality of Life: Expanding a Conceptual Framework to Include Older Adults Who Receive Long-term Services and Supports. The Gerontologist. 2013 Apr 1;53(2):205–10.

10. Paterson DH, Jones GR, Rice CL. Ageing and physical activity: evidence to develop exercise recommendations for older adults. Can J Public Health. 2007;98 Suppl 2:S69–108.

11. Paterson DH, Warburton DE. Physical activity and functional limitations in older adults: a systematic review related to Canada’s Physical Activity Guidelines. Int J Behav Nutr Phys Act. 2010;7(1):38.

12. Gauthier AH, Smeeding TM. Time Use at Older Ages Cross-National Differences. Res Aging. 2003 May 1;25(3):247–74.

13. Tudor-Locke C, Johnson WD, Katzmarzyk PT. Frequently reported activities by intensity for U.S. adults: the American Time Use Survey. Am J Prev Med. 2010 Oct;39(4):e13–20.

14. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires * Commentary. Br J Sports Med. 2003 Jun;37(3):197–206.

15. Bonnefoy M, Normand S, Pachiaudi C, Lacour JR, Laville M, Kostka T. Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. J Am Geriatr Soc. 2001 Jan;49(1):28–35.

16. Baranowski T. Validity and Reliability of Self Report Measures of Physical Activity: An Information-Processing Perspective. Res Q Exerc Sport. 1988;59(4):314–27.

17. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Kokmen E, Tangelos EG. Aging, memory, and mild cognitive impairment. Int Psychogeriatr. 1997;9(S1):65–9.

18. Washburn RA, Jette AM, Janney CA. Using Age-Neutral Physical Activity Questionnaires in Research with the Elderly. J Aging Health. 1990;2(3):341 –356.

19. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008 Jan;40(1):181–8.

20. Davis MG, Fox KR. Physical activity patterns assessed by accelerometry in older people. Eur J Appl Physiol. 2007 Jul;100(5):581–9.

21. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. J Aging Phys Act. 2009 Jan;17(1):17–30.

22. Miller NE, Strath SJ, Swartz AM, Cashin SE. Estimating absolute and relative physical activity intensity across age via accelerometry in adults. J Aging Phys Act. 2010 Apr;18(2):158–70.

23. Davis MG, Fox KR, Hillsdon M, Sharp DJ, Coulson JC, Thompson JL. Objectively measured physical activity in a diverse sample of older urban UK adults. Med Sci Sports Exerc. 2011 Apr;43(4):647–54.

24. Matthew CE. Calibration of Accelerometer Output for Adults. Med Sci Sports Exerc. 2005 Nov;37(Supplement):S512–S522.

25. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. Pervasive Computing. 2004;1–17.

26. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol. 2009 Jul;107(4):1300–7.

27. De Vries SI, Garre FG, Engbers LH, Hildebrandt VH, Van Buuren S. Evaluation of Neural Networks to Identify Types of Activity Using Accelerometers. Med Sci Sports Exerc. 2011 Jan;43(1):101–7.

28. Kohl HW 3rd. Physical activity and cardiovascular disease: evidence for a dose response. Med Sci Sports Exerc. 2001 Jun;33(6 Suppl):S472–483; S493–494.

29. Manini TM, Everhart JE, Patel KV, Schoeller DA, Cummings S, Mackey DC, et al. Activity Energy Expenditure and Mobility Limitation in Older Adults: Differential Associations by Sex. Am J Epidemiol. 2009 Jun 15;169(12):1507–16.

30. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. Can Med Assoc J. 2006 Mar 14;174(6):801–9.

31. Manini TM, Everhart JE, Patel KV, et al. DAily activity energy expenditure and mortality among older adults. J Am Med Assoc. 2006 Jul 12;296(2):171–9.

32. Preece SJ, Goulermas JY, Kenney LPJ, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. Physiol Meas. 2009 Apr;30(4):R1–33.

33. Visser M, Simonsick EM, Colbert LH, Brach J, Rubin SM, Kritchevsky SB, et al. Type and Intensity of Activity and Risk of Mobility Limitation: The Mediating Role of Muscle Parameters. J Am Geriatr Soc. 2005 May;53(5):762–70.

34. Yang Y-R, Wang R-Y, Lin K-H, Chu M-Y, Chan R-C. Task-oriented progressive resistance strength training improves muscle strength and functional performance in individuals with stroke. Clin Rehabil. 2006 Oct;20(10):860–70.

35. Alexander NB, Galecki AT, Grenier ML, Nyquist LV, Hofmeyer MR, Grunawalt JC, et al. Task-Specific Resistance Training to Improve the Ability of Activities of Daily Living–Impaired Older Adults to Rise from a Bed and from a Chair. J Am Geriatr Soc. 2001 Nov 1;49(11):1418–27.

36. Brownson RC, Boehmer TK, Luke DA. Declining Rates of Physical Activity in The United States: What Are the Contributors? Annu Rev Public Health. 2005 Apr;26(1):421–43.

37. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol. 2008 Apr 1;167(7):875–81.

38. Healy GN, Dunstan DW, Salmon J, Cerin E, Shaw JE, Zimmet PZ, et al. Breaks in sedentary time: beneficial associations with metabolic risk. Diabetes Care. 2008 Apr;31(4):661–6.

39. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc. 2011 Aug;43(8):1561–7.

40. Tapia EM, Intille SS, Haskell W, Larson K, Wright J, King A, et al. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. 2007;

41. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. Med Sci Sports Exerc. 2012 Apr;44(4):742–8.

42. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. J Appl Physiol. 2011 Dec 1;111(6):1804–12.

43. Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. J Appl Physiol. 2007 Oct;103(4):1419–27.

44. Najafi B, Aminian K, Paraschiv-Ionescu A, Loew F, Bula CJ, Robert P. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. IEEE Trans Biomed Eng. 2003 Jun;50(6):711–23.

45. Bourke AK, O'brien JV, Lyons GM. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. Gait Posture. 2007;26(2):194–9.

46. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Br Med J. 1992 Jul 18;305(6846):160–4.

47. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol. 1994 Mar;49(2):M85–94.

48. Rikli R, Jones C. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. J Aging Phys Activ. 1999;7:129–61.

49. Cress ME, Buchner DM, Questad KA, Esselman PC, deLateur BJ, Schwartz RS. Continuous-scale physical functional performance in healthy older adults: a validation study. Arch Phys Med Rehabil. 1996 Dec;77(12):1243–50.

50. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. Med Care. 1992 Jun 1;30(6):473–83.

51. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med. 1995;332(9):556–62.

52. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. Eur J Appl Physiol. 2010 Sep;111(2):187–201.

53. Crouter SE, Churilla JR, Bassett DR Jr. Estimating energy expenditure using accelerometers. Eur J Appl Physiol. 2006 Dec;98(6):601–12.

54. Studenski S, Perera S, Patel K, et al. GAit speed and survival in older adults. J Am Med Assoc. 2011 Jan 5;305(1):50–8.

55. Kim M-J, Yabushita N, Kim M-K, Nemoto M, Seino S, Tanaka K. Mobility performance tests for discriminating high risk of frailty in community-dwelling older women. Arch Gerontol Geriatr. 2010 Oct;51(2):192–8.

56. Song Y, Shin S, Kim S, Lee D, Lee KH. Speed estimation from a tri-axial accelerometer using neural networks. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:3224–7.

57. Mannini A, Sabatini AM. On-line classification of human activity and estimation of walk-run speed from acceleration data using support vector machines. Conf Proc IEEE Eng Med Biol Soc. 2011;2011:3302–5.

58. Carlson RH Jr, Huebner DR, Hoarty CA, Whittington J, Haynatzki G, Balas MC, et al. Treadmill gait speeds correlate with physical activity counts measured by cell phone accelerometers. Gait Posture. 2012 Jun;36(2):241–8.

59. Mannini A, Intille SS, Rosenberger M, Sabatini AM, Haskell W. Activity recognition using a single accelerometer placed at the wrist or ankle. Med Sci Sports Exerc. 2013 Nov;45(11):2193–203.

60. Potter JM, Evans AL, Duncan G. Gait speed and activities of daily living function in geriatric patients. Arch Phys Med Rehabil. 1995 Nov;76(11):997–9.

61. Huang W-NW, Perera S, VanSwearingen J, Studenski S. Performance Measures Predict the Onset of Basic ADL Difficulty in Community-Dwelling Older Adults. J Am Geriatr Soc. 2010 May;58(5):844–52.

62. Foerster F, Smeja M, Fahrenberg J. Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Comput Hum Behav. 1999 Sep 1;15(5):571–83.

63. Ermes M, Parkka J, Mantyjarvi J, Korhonen I. Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions. IEEE Trans Inf Technol Biomed. 2008;12(1):20–6.

64. Gyllensten IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. IEEE Trans Biomed Eng. 2011 Sep;58(9):2656–63.

65. Swartz A, Strath SJ, Bassett DR, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc. 2000;32(9):S450.

66. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. Med Sci Sports Exerc. 2000;32(9):S442.

67. Bassett DR. Validity of four motion sensors in measuring moderate intensity physical activity. Med Sci Sports Exerc. 2000;32(9):S471.

68. Leenders NY, Sherman WM, Nagaraja H, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. Med Sci Sports Exerc. 2001;33(7):1233.

69. Waldrop J, Stern SM, Bureau UC. Disability Status, 2000 [Internet]. US Dept. of Commerce, Economics and Statistics Administration, US Census Bureau; 2003 [cited 2011 Jun 20]. Available from: http://www.census.gov/prod/2003pubs/c2kbr-17.pdf

70. Congressional Budget Office. Projections of Expenditures for Long-Term Care Services for the Elderly. 1999 Mar [cited 2011 Aug 4]; Available from: http://www.cbo.gov

71. Boyle PA, Buchman AS, Wilson RS, Bienias JL, Bennett DA. Physical Activity Is Associated with Incident Disability in Community-Based Older Persons. J Am Geriatr Soc. 2007 Feb;55(2):195–201.

72. Leveille SG, Guralnik JM, Ferrucci L, Langlois JA. Aging successfully until death in old age: opportunities for increasing active life expectancy. Am J Epidemiol. 1999;149(7):654.

73. Miller ME, Rejeski WJ, Reboussin BA, Ten Have TR, Ettinger WH. Physical activity, functional limitations, and disability in older adults. J Am Geriatr Soc. 2000 Oct;48(10):1264–72.

74. Proctor DN, Melton Iii LJ, Khosla S, Crowson CS, O'Connor MK, Riggs BL. Relative influence of physical activity, muscle mass and strength on bone density. Osteoporosis int. 2000;11(11):944–52.

75. Rantanen T, Guralnik JM, Sakari-Rantala R, Leveille S, Simonsick EM, Ling S, et al. Disability, physical activity, and muscle strength in older women: The women's health and aging study* 1. Arch Phys Med Rehabil. 1999;80(2):130–5.

76. Kohrt WM, Bloomfield SA, Little KD, Nelson ME, Yingling VR. Physical Activity and Bone Health. Med Sci Sports Exerc. 2004 Nov;36(11):1985–96.

77. Wilson TM, Tanaka H. Meta-analysis of the age-associated decline in maximal aerobic capacity in men. J Appl Physiol. 2002;92(6):2303–8.

78. Tanaka H, Desouza CA, Jones PP, Stevenson ET, Davy KP, Seals DR. Greater rate of decline in maximal aerobic capacity with age in physically active vs. sedentary healthy women. J Appl Physiol. 1997;83(6):1947.

79. Manini TM, Pahor M. Physical activity and maintaining physical function in older adults. Br J Sports Med. 2008 Nov;43(1):28–31.

80. Brach JS, Simonsick EM, Kritchevsky S, Yaffe K, Newman AB. The association between physical function and lifestyle activity and exercise in the health, aging and body composition study. J Am Geriatr Soc. 2004 Apr;52(4):502–9.

81. Chalé-Rush A, Guralnik JM, Walkup MP, Miller ME, Rejeski WJ, Katula JA, et al. Relationship Between Physical Functioning and Physical Activity in the Lifestyle Interventions and Independence for Elders Pilot. J Am Geriatr Soc. 2010 Oct;58(10):1918–24.

82. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008;5(1):56.

83. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol. 2008 Apr 1;167(7):875–81.

84. Dunstan DW, Barr ELM, Healy GN, Salmon J, Shaw JE, Balkau B, et al. Television viewing time and mortality: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Circulation. 2010 Jan 26;121(3):384–91.

85. Gardiner PA, Healy GN, Eakin EG, Clark BK, Dunstan DW, Shaw JE, et al. Associations between television viewing time and overall sitting time with the metabolic syndrome in older men and women: the Australian diabetes obesity and lifestyle study. J Am Geriatr Soc. 2011 May;59(5):788–96.

86. Thorp AA, Healy GN, Owen N, Salmon J, Ball K, Shaw JE, et al. Deleterious associations of sitting time and television viewing time with cardiometabolic risk biomarkers: Australian Diabetes, Obesity and Lifestyle (AusDiab) study 2004-2005. Diabetes Care. 2010 Feb;33(2):327–34.

87. Wijndaele K, Healy GN, Dunstan DW, Barnett AG, Salmon J, Shaw JE, et al. Increased cardiometabolic risk is associated with increased TV viewing time. Med Sci Sports Exerc. 2010 Aug;42(8):1511–8.

88. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc. 2009 May;41(5):998–1005.

89. Dipietro L, Caspersen CJ, Ostfeld AM, Nadel ER. A survey for assessing physical activity among older adults. Med Sci Sports Exerc. 1993 May;25(5):628–42.

90. Freedson PS, Miller K. Objective monitoring of physical activity using motion sensors and heart rate. Res Q Exerc Sport. 2000 Jun;71(2 Suppl):S21–29.

91. Schneider PL, Crouter SE, Bassett DR. Pedometer Measures of Free-Living Physical Activity: Comparison of 13 Models. Med Sci Sports Exerc. 2004 Feb;36(2):331–5.

92. Berlin JE, Storti KL, Brach JS. Using Activity Monitors to Measure Physical Activity in Free-Living Conditions. Physical Therapy. 2006 Aug;86(8):1137 –1145.

93. Melanson EL, Knoll JR, Bell ML, Donahoo WT, Hill JO, Nysse LJ, et al. Commercially available pedometers: considerations for accurate step counting. Prev Med. 2004 Aug;39(2):361–8.

94. Leenders NY, Sherman WM, Nagaraja HN, Kien CL. Evaluation of methods to assess physical activity in free-living conditions. Med Sci Sports Exerc. 2001 Jul;33(7):1233–40.

95. Ewald B, McEvoy M, Attia J. Pedometer counts superior to physical activity scale for identifying health markers in older adults. Br J Sports Med. 2010 Aug;44(10):756–61.

96. Melanson EL, Knoll JR, Bell ML, Donahoo WT, Hill JO, Nysse LJ, et al. Commercially available pedometers: considerations for accurate step counting. Prev Med. 2004 Aug;39(2):361–8.

97. Storti KL, Pettee KK, Brach JS, Talkowski JB, Richardson CR, Kriska AM. Gait speed and step-count monitor accuracy in community-dwelling older adults. Med Sci Sports Exerc. 2008 Jan;40(1):59–64.

98. Le Masurier GC, Tudor-Locke C. Comparison of pedometer and accelerometer accuracy under controlled conditions. Med Sci Sports Exerc. 2003 May;35(5):867–71.

99. Dijkstra B, Zijlstra W, Scherder E, Kamsma Y. Detection of walking periods and number of steps in older adults and patients with Parkinson's disease: accuracy of a pedometer and an accelerometry-based method. Age Ageing. 2008 Jul;37(4):436–41.

100. Chen KY, Bassett DR Jr. The technology of accelerometry-based activity monitors: current and future. Med Sci Sports Exerc. 2005 Nov;37(11 Suppl):S490–500.

101. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc. 1998 May;30(5):777–81.

102. Crouter SE. A novel method for using accelerometer data to predict energy expenditure. J Appl Physiol. 2005 Dec;100(4):1324–31.

103. Marshall AL, Miller YD, Burton NW, Brown WJ. Measuring total and domain-specific sitting: a study of reliability and validity. Med Sci Sports Exerc. 2010 Jun;42(6):1094–102.

104. Rosenberg DE, Bull FC, Marshall AL, Sallis JF, Bauman AE. Assessment of sedentary behavior with the International Physical Activity Questionnaire. J Phys Act Health. 2008;5 Suppl 1:S30–44.

105. Langley P. Machine Learning as an Experimental Science. Machine Learning. 1988;3(1):5–8.

106. Bishop CM. Pattern recognition and machine learning. Springer; 2009.

107. Fahrenberg J, Foerster F, Smeja M, Müller W. Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. Psychophysiology. 1997 Sep;34(5):607–12.

108. Fahrenberg J, Müller W, Foerster F, Smeja M. A multi-channel investigation of physical activity. J Psychophysiol. 1996;10:209–17.

109. Cheung VH, Gray L, Karunanithi M. Review of accelerometry for determining daily activity among elderly patients. Arch Phys Med Rehabil. 2011 Jun;92(6):998–1014.

110. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENEA: optimum sampling frequency and number of axes. Med Sci Sports Exerc. 2012 Nov;44(11):2228–34.

111. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. Med Sci Sports Exerc. 2006 Sep;38(9):1626–34.

112. Culhane KM, Lyons GM, Hilton D, Grace PA, Lyons D. Long-term mobility monitoring of older adults using accelerometers in a clinical environment. Clin Rehabil. 2004 May;18(3):335–43.

113. Vestergaard S, Patel KV, Bandinelli S, Ferrucci L, Guralnik JM. Characteristics of 400-meter walk test performance and subsequent mortality in older adults. Rejuvenation Res. 2009 Jun;12(3):177–84.

114. Newman AB, Simonsick EM, Naydeck BL, Boudreau RM, Kritchevsky SB, Nevitt MC, et al. Association of Long-Distance Corridor Walk Performance With Mortality, Cardiovascular Disease, Mobility Limitation, and Disability. J Am Med Assoc. 2006 May 3;295(17):2018–26.

115. Cavanaugh JT, Kochi N, Stergiou N. Nonlinear analysis of ambulatory activity patterns in community-dwelling older adults. J Gerontol A Biol Sci Med Sci. 2010 Feb;65(2):197–203.

116. Foerster F, Fahrenberg J. Motion pattern and posture: correctly assessed by calibrated accelerometers. Behav Res Methods Instrum Comput. 2000 Aug;32(3):450–7.

117. Yang C-C, Hsu Y-L. Development of a wearable motion detector for telemonitoring and real-time identification of physical activity. Telemed J E Health. 2009 Jan;15(1):62–72.

118. Hughes VA, Frontera WR, Wood M, Evans WJ, Dallal GE, Roubenoff R, et al. Longitudinal muscle strength changes in older adults. J Geront A Biol Sci Med Sci. 2001;56(5):B209.

119. Fleg JL, Morrell CH, Bos AG, Brant LJ, Talbot LA, Wright JG, et al. Accelerated Longitudinal Decline of Aerobic Capacity in Healthy Older Adults. Circulation. 2005 Aug 2;112(5):674–82.

120. Kozey SL, Lyden K, Howe CA, Staudenmayer JW, Freedson PS. Accelerometer output and MET values of common physical activities. Med Sci Sports Exerc. 2010 Sep;42(9):1776–84.

121. Gunn SM, Brooks AG, Withers RT, Gore CJ, Plummer JL, Cormack J. The energy cost of household and garden activities in 55- to 65-year-old males. Eur J Appl Physiol. 2005 Jul;94(4):476–86.

122. Cooper JA, Watras AC, O'Brien MJ, Luke A, Dobratz JR, Earthman CP, et al. Assessing validity and reliability of Resting Metabolic Rate in six gas analysis systems. J Am Diet Assoc. 2009 Jan;109(1):128–32.

123. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. Eur J Appl Physiol. 2009 Dec;109(2):159–71.

124. Shigematsu R, Ueno LM, Nakagaichi M, Nho H, Tanaka K. Rate of perceived exertion as a tool to monitor cycling exercise intensity in older adults. J Aging Phys Act. 2004 Jan;12(1):3–9.

125. Wallerstein LF, Tricoli V, Barroso R, Rodacki A LF, Russo L, Aihara AY, et al. Effects of strength and power training on neuromuscular variables in older adults. J Aging Phys Act. 2012 Apr;20(2):171–85.

126. Seco J, Abecia LC, Echevarría E, Barbero I, Torres-Unda J, Rodriguez V, et al. A long-term physical activity training program increases strength and flexibility, and improves balance in older adults. Rehabil Nurs. 2013 Feb;38(1):37–47.

127. Mackay-Lyons M. Aerobic treadmill training effectively enhances cardiovascular fitness and gait function for older persons with chronic stroke. J Physiother. 2012;58(4):271.

128. Bonomi AG, Goris AHC, Yin B, Westerterp KR. Detection of type, duration, and intensity of physical activity using an accelerometer. Med Sci Sports Exerc. 2009 Sep;41(9):1770–7.

129. NHANES - NHANES 2011-2012 - Manuals, Brochures, and Consent Documents [Internet]. [cited 2013 Oct 23]. Available from: http://www.cdc.gov/nchs/nhanes/nhanes2011-2012/current_nhanes_11_12.htm

130. Barstow TJ, Molé PA. Linear and nonlinear characteristics of oxygen uptake kinetics during heavy exercise. J Appl Physiol. 1991 Dec;71(6):2099–106.

131. Takezawa K. Linear Mixed Model. Learning Regression Analysis by Simulation [Internet]. Springer Japan; 2014 [cited 2013 Nov 8]. p. 269–94. Available from: http://link.springer.com/chapter/10.1007/978-4-431-54321-3_6

132. Jackson AS, Blair SN, Mahar MT, Wier LT, Ross RM, Stuteville JE. Prediction of functional aerobic capacity without exercise testing. Med Sci Sports Exerc. 1990 Dec;22(6):863–70.

133. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. A Method to Estimate Free-Living Active and Sedentary Behavior from an Accelerometer. Med Sci Sports Exerc. 2013 Jul 15;

134. Stamatakis E, Davis M, Stathi A, Hamer M. Associations between multiple indicators of objectively-measured and self-reported sedentary behaviour and cardiometabolic risk in older adults. Prev Med. 2012 Jan;54(1):82–7.

135. Santos DA, Silva AM, Baptista F, Santos R, Vale S, Mota J, et al. Sedentary behavior and physical activity are independently related to functional fitness in older adults. Exp Gerontol. 2012 Dec;47(12):908–12.

136. Kozey S, Lyden K, Staudenmayer J, Freedson P. Errors in MET estimates of physical activities using 3.5 ml x kg(-1) x min(-1) as the baseline oxygen consumption. J Phys Act Health. 2010 Jul;7(4):508–16.

137. Hall KS, Howe CA, Rana SR, Martin CL, Morey MC. METs and accelerometry of walking in older adults: standard versus measured energy cost. Med Sci Sports Exerc. 2013 Mar;45(3):574–82.

138. John D, Liu S, Sasaki JE, Howe CA, Staudenmayer J, Gao RX, et al. Calibrating a novel multi-sensor physical activity measurement system. Physiol Meas. 2011 Sep;32(9):1473–89.

139. Breiman L. Random Forests. Machine Learning. 2001 Oct 1;45(1):5–32.

140. Physical Activity for Everyone: Guidelines: Adults | DNPAO | CDC [Internet]. [cited 2012 Oct 29]. Available from: http://www.cdc.gov/physicalactivity/everyone/guidelines/adults.html

141. Bruce DG, Devine A, Prince RL. Recreational Physical Activity Levels in Healthy Older Women: The Importance of Fear of Falling. J Am Geriatr Soc. 2002;50(1):84–9.

142. Rogers ME, Rogers NL, Takeshima N, Islam MM. Methods to assess and improve the physical parameters associated with fall risk in older adults. Prev Med. 2003 Mar;36(3):255–64.